



**DEPARTAMENTO DE BIOLOGÍA MOLECULAR**

**CARACTERIZACIÓN MOLECULAR DEL  
BACTERIÓFAGO BAM35. MECANISMO DE  
REPLICACIÓN DEL DNA Y ESTUDIO DEL  
INTERACTOMA PROTEICO**

**TESIS DOCTORAL**

**Monica Rosa Berjón Otero**

**Madrid, 2017**



**Departamento de Biología Molecular  
Facultad de Ciencias  
Universidad Autónoma de Madrid**

# **CARACTERIZACIÓN MOLECULAR DEL BACTERIÓFAGO BAM35. MECANISMO DE REPLICACIÓN DEL DNA Y ESTUDIO DEL INTERACTOMA PROTEICO**

Memoria de investigación presentada por la Licenciada en Biotecnología  
**Mónica Rosa Berjón Otero** para optar al grado de Doctora en Biología  
Molecular por la Universidad Autónoma de Madrid.





Esta tesis doctoral ha sido realizada en el Centro de Biología Molecular Severo Ochoa (CSIC/UAM) bajo la dirección de los Doctores Margarita Salas Falgueras y Modesto Redrejo Rodríguez. Para su realización se contó con una beca de Formación de Personal Investigador del Ministerio de Economía, Industria y Competitividad.



*Los imposibles de hoy serán posibles mañana*  
Konstantin Tsiolkovsky



A mis padres, Mónica y Carlos



|  |           |
|--|-----------|
| <b>Abreviaturas.....</b>   | <b>1</b>  |
| <b>Glosario de términos no castellanos incluidos en esta Tesis .....</b>   | <b>5</b>  |
| <b>Resumen .....</b>   | <b>7</b>  |
| <b>Summary .....</b>   | <b>9</b>  |
| <b>Introducción.....</b>   | <b>11</b> |
| 1. Familia <i>Tectiviridae</i> : características generales.....  | 14        |
| 2. Relación filogenética de los tectivirus con otros grupos de virus.....  | 15        |
| 3. El bacteriófago Bam35 como modelo de estudio.....   | 17        |
| 3.1. Morfología de la partícula viral.....   | 17        |
| 3.2. Genoma de Bam35 .....   | 20        |
| 4. Técnicas de estudio de la complejidad de los organismos y su aplicación en bacteriófagos .....                                | 21        |
| 5. DNA polimerasas.....  | 23        |
| 5.1. Familias de DNA polimerasas .....   | 23        |
| 5.2. Estructura de las DNA polimerasas .....   | 26        |
| 6. Replicación iniciada por proteína terminal .....  | 27        |
| 6.1. Mecanismos de recuperación de la información genética de los extremos 3' del genoma en la replicación iniciada por TP ..... | 29        |
| 7. Fidelidad en la replicación del DNA.....  | 31        |
| 7.1. Síntesis de DNA a través de lesiones .....  | 32        |
| <b>Objetivos.....</b>  | <b>35</b> |
| <b>Materiales y Métodos .....</b>  | <b>39</b> |
| 1. Nucleótidos.....  | 41        |
| 2. Oligonucleótidos y DNA .....  | 41        |
| 3. Proteínas.....  | 41        |
| 4. DNA polimerasa <i>wild type</i> de Bam35 y variantes .....  | 41        |
| 4.1. Obtención de vectores de expresión.....   | 41        |
| 4.2. Expresión y purificación.....   | 42        |
| 5. Proteína terminal <i>wild type</i> de Bam35 y variantes .....   | 43        |
| 5.1. Obtención de vectores de expresión.....   | 43        |
| 5.2. Expresión y purificación.....   | 44        |
| 6. Obtención de TP-DNA de Bam35 .....  | 45        |
| 7. Ensayos con la TP y la B35DNAP .....  | 46        |
| 7.1. Ensayo de iniciación (formación del complejo TP-dNMP) .....   | 46        |

|   |           |
|---|-----------|
| 7.2. Mapeo del residuo iniciador de la TP de Bam35.....   | 47        |
| 7.3. Ensayo de iniciación con extractos bacterianos.....  | 47        |
| 7.4. Análisis de la interacción TP/B35DNAP a través de un ensayo funcional.....   | 48        |
| 7.5. Replicación trunca y completa de oligonucleótidos de banda simple .....  | 48        |
| 7.6. Replicación de TP-DNA .....  | 48        |
| 8. Ensayos con la B35DNAP.....  | 49        |
| 8.1. Replicación y degradación de oligonucleótidos .....  | 49        |
| 8.2. Retraso en gel .....   | 50        |
| 8.3. Replicación del DNA circular de M13.....   | 50        |
| 8.4. Amplificación por desplazamiento múltiple (MDA) de un plásmido de DNA.....   | 51        |
| 8.5. Análisis de la fidelidad de la B35DNAP .....   | 51        |
| 9. Sistema de doble híbrido de levaduras.....   | 52        |
| 9.1. Clonaje de los ORFs de Bam35 en el vector de entrada pDONR/Zeo.....  | 52        |
| 9.2. Obtención de los vectores de expresión en levaduras .....  | 55        |
| 9.3. Sistema de doble híbrido en levaduras .....  | 56        |
| <b>Resultados.....</b>  | <b>57</b> |
| <b>Capítulo 1. Caracterización bioquímica inicial de la DNA polimerasa de Bam35.....</b>  | <b>59</b> |
| 1. Comparación de las secuencias aminoácidicas de la B35DNAP y la $\Phi$ 29DNAP .....   | 61        |
| 2. Optimización del sustrato para ensayos de replicación de oligonucleótidos por la B35DNAP .....   | 62        |
| 3. Las actividades de polimerización y exonucleasa de la B35DNAP están coordinadas .....  | 65        |
| 4. La B35DNAP sintetiza el DNA de manera procesiva acoplada al desplazamiento de banda .....  | 67        |
| 5. La B35DNAP amplifica el DNA con alta fidelidad siendo su principal error la inserción de bases en secuencias homopoliméricas .....               | 69        |
| <b>Capítulo 2. Estudio de las primeras etapas de replicación del DNA de Bam35 .....</b>   | <b>73</b> |
| 1. El ORF4 del genoma de Bam35 codifica una TP funcional .....  | 75        |
| 2. La tirosina conservada 194 de la TP de Bam35 es el residuo iniciador de la replicación del DNA.....  | 78        |
| 3. La tirosina conservada 172 de la TP de Bam35 no es necesaria para la iniciación pero está implicada en la interacción con la B35DNAP .....       | 82        |
| 4. La iniciación de la replicación del genoma de Bam35 está dirigida principalmente por el tercer nucleótido del extremo 3' de la cadena molde..... | 83        |
| 5. La información del extremo del genoma de Bam35 puede ser recuperada mediante un nuevo mecanismo de <i>jumping-back</i> .....                     | 86        |



|   |            |
|---|------------|
| <b>Capítulo 3. Estudio de la capacidad de replicación a través de daños de la DNA polimerasa de Bam35 .....</b>   | <b>89</b>  |
| 1. La B35DNAP es capaz de replicar un molde de DNA con sitios abásicos insertando preferentemente una adenina enfrente del daño .....   | 91         |
| 2. La secuencia de DNA influye en la eficacia de inserción del nucleótido enfrente del sitio abásico y en la elongación posterior del <i>primer</i> .....   | 94         |
| 3. La actividad correctora de pruebas de la B35DNAP contrarresta la replicación de oligonucleótidos con sitios abásicos hasta que el daño se encuentra a más de 5 nucleótidos de distancia del extremo 3' del <i>primer</i> ..... | 96         |
| 4. La eliminación del subdomino TPR2 de la B35DNAP no afecta a la insercción del nucleótido enfrente del THF, pero impide la extensión posterior del <i>primer</i> .....  | 99         |
| <b>Capítulo 4. Mapeo de las interacciones entre las proteínas codificadas por el genoma de Bam35.....</b>   | <b>103</b> |
| 1. Obtención de librerías de expresión de todos los ORFs de Bam35 en cuatro vectores de levaduras.....  | 105        |
| 2. Determinación de las condiciones óptimas de selección .....  | 107        |
| 3. Interactoma proteico de Bam35.....   | 108        |
| 3.1. Establecimiento de los criterios seguidos para la interpretación de los resultados. Eliminación de los posibles falsos positivos.....  | 108        |
| 3.2. Interacciones detectadas con la combinación de los vectores pGBKCg y pGADCg (CC) .....   | 111        |
| 3.3. Interacciones detectadas con la combinación de los vectores pGBKCg y pGADT7g (CN) .....  | 112        |
| 3.4. Interacciones detectadas con la combinación de los vectores pGBGT7g y pGADT7g (NN).....  | 113        |
| 3.5. Interacciones detectadas con la combinación de los vectores pGBGT7g y pGADCg (NC) .....  | 114        |
| 4. Interactoma proteico de Bam35.....   | 115        |
| <b>Discusión .....</b>  | <b>117</b> |
| 1. La B35DNAP es una enzima procesiva y fiel .....  | 119        |
| 2. Iniciación de la replicación del genoma de Bam35 mediada por TP .....  | 120        |
| 2.1. Deoxinucleotidilación de la TP por la B35DNAP.....   | 120        |
| 2.2. La replicación del tamaño completo del TP-DNA <i>in vitro</i> requiere solo de dos proteínas.....  | 121        |
| 2.3. Los residuos conservados tirosina 194 y 172 de la TP son esenciales para la iniciación.....  | 122        |
| 2.4. Un nuevo mecanismo de recuperación de la información genética del extremo 3' del genoma: <i>Jumping-back</i> de un nucleótido .....  | 123        |
| 3. La B35DNAP: Una enzima con alta fidelidad capaz de replicar DNAs con sitios abásicos.....  | 124        |
| 3.1. La "Memoria a corto plazo" de la capacidad correctora de pruebas contrarresta la replicación de DNA dañado .....   | 126        |
| 4. La red de interacción de las proteínas de Bam35 revela el papel de nuevas proteínas estructurales.....   | 127        |

|  |            |
|--|------------|
| 4.1. Proteínas de la cápsida y de la espícula .....                      | 128        |
| 4.2. Proteínas de membrana.....  | 129        |
| 4.3. Proteínas de replicación y regulación del ciclo.....                | 130        |
| 5. Conclusión: relevancia de los resultados y perspectivas futuras ..... | 132        |
| <b>Conclusiones .....</b>  | <b>133</b> |
| <b>Conclusions .....</b>   | <b>137</b> |
| <b>Bibliografía .....</b>  | <b>141</b> |
| <b>Material suplementario.....</b>                                       | <b>165</b> |

**aa:** aminoácidos

**AD:** dominio de activación del factor de transcripción eucariótico Gal4 (del inglés, *Activation Domain*)

**Amp<sup>R</sup>:** resistencia al antibiótico ampicilina

**BSA:** albúmina de suero bovino (del inglés, *Bovine Serum Albumin*)

**B35DNAP:** DNA polimerasa del bacteriófago Bam35

**C-terminal:** carboxi-terminal

**CNBr:** bromuro de cianógeno

**dAMP:** 2'-desoxiadenosina-5'-monofosfato

**dATP:** 2'-desoxiadenosina-5'-trifosfato

**DBD:** dominio de unión al DNA del factor de transcripción eucariótico Gal4 (del inglés, *DNA-Binding Domain*)

**DBP:** proteína de unión a DNA de banda doble (del inglés, *Double-stranded Binding Protein*)

**DE:** desviación estándar

**dCMP:** 2'-desoxicitidina-5'-monofosfato

**dCTP:** 2'-desoxicitidina-5'-trifosfato

**ddA:** 2', 3'-didesoxiadenosina-5'-monofosfato

**ddATP:** 2', 3'-didesoxiadenosina-5'-trifosfato

**ddNTP:** 2', 3'-didesoxinucleósido-5'-trifosfato

**ddT:** 2', 3'-didesoxitimidina-5'-monofosfato

**ddTTP:** 2', 3'-didesoxitimidina-5'-trifosfato

**dGMP:** 2'-desoxiguanosina-5'-monofosfato

**dGTP:** 2'-desoxiguanosina-5'-trifosfato

**DNA:** ácido desoxirribonucleico (del inglés, *DeoxyriboNucleic Acid*)

**DNAP:** DNA polimerasa

**dNMP:** 2'-desoxinucleósido-5'-monofosfato

**dNTP:** 2'-desoxinucleósido-5'-trifosfato

**DO<sub>600</sub>:** densidad óptica medida a una longitud de onda de 600 nm

**dsDNA:** ácido desoxirribonucleico de doble cadena (del inglés, *Double-Stranded DNA*)

**dTMP:** 2'-desoxitimidina-5'-monofosfato

**DTT:** ditioneitol

**dTTP:** 2'-desoxitimidina-5'-trifosfato

**EDTA:** ácido etilendiaminotetraacético (del inglés, *EthyleneDiamineTetraacetic Acid*)

**Gen<sup>R</sup>:** resistencia a gentamicina

**IPTG:** isopropil β-D-tiogalactopiranosido

**ITRs:** repeticiones terminales invertidas (del inglés, *Inverted Terminal Repeat*)

**Kan<sup>R</sup>**: resistencia a kanamicina

**Kb**: Kilo base

**KDa**: Kilo Dalton

**Kpb**: Kilo pares de bases

**LB**: medio de cultivo Luria Bertani

**Mb**: Mega base

**MCP**: proteína mayoritaria de la cápsida (del inglés, *Majority Capsid Protein*)

**MDA**: amplificación por desplazamiento múltiple (del inglés, *Multiple Displacement Amplification*)

**mer**: número de nucleótidos de un oligómero

**min**: minutos

**MMR**: mecanismo de reparación de errores producidos durante la replicación del DNA (del inglés, *Mismatch Repair*)

**N-terminal**: amino-terminal

**OH**: hidroxilo

**ORF**: marco de lectura abierto (del inglés, *Open Reading Frame*)

**PAGE**: electroforesis en gel de poliacrilamida (del inglés, *PolyAcrylamide Gel Electrophoresis*)

**pb**: par de bases

**PCR**: reacción en cadena de la polimerasa (del inglés, *Polymerase Chain Reaction*)

**pI**: punto isoeléctrico

**rNTP**: ribonucleósido 5'-trifosfato

**SDS**: dodecilsulfato sódico (del inglés, *Sodium Dodecyl Sulfate*)

**SSB**: proteína de unión a DNA de cadena sencilla (del inglés, *Single-Stranded DNA Binding protein*)

**ssDNA**: DNA de cadena sencilla (del inglés, *Single-Stranded DNA*)

**THF**: tetrahidrofurano

**TLS**: síntesis a través de una lesión en el DNA (del inglés, *Translesion Synthesis*)

**Tm**: temperatura de fusión (del inglés, *Melting Temperature*)

**TP**: proteína terminal (del inglés, *Terminal Protein*)

**TP-dAMP, TP-dTMP, TP-dGMP, TP-dCMP**: complejo covalente formado entre la proteína terminal y el nucleótido indicado

**TP-DNA**: DNA con una proteína terminal covalentemente unida a los extremos 5'

**TP-dNMP**: complejo covalente formado entre la proteína terminal y un dNMP

**TPR1 y TPR2**: subdominios característicos de las DNA polimerasas que inician la replicación utilizando una proteína como *primer* (del inglés, *Terminal Protein Region*)

**Tris**: tri(hidroximetil) aminometano

**UAS:** secuencia de activación localizada *upstream* del promotor (del inglés *Upstream Activating Sequence*)

**ufp:** unidad formadora de placa

**WT:** tipo natural o silvestre, sin modificación genética (del inglés, *Wild Type*)

**x-Gal:** 5-bromo-4-cloro-3-indolil- $\beta$ -D-galactopiranosido

**YFP:** proteína amarilla fluorescente (del inglés, *Yellow Fluorescent Protein*)

**YFPTP:** proteína de fusión de la YFP al N-terminal de la proteína terminal de Bam35

**Zeo<sup>R</sup>:** resistencia al antibiótico Zeocina

**3AT:** 3-amino-1,2,4-triazol

**8-oxoG:** 7, 8-dihidro-8-oxo-2'-oxodeoxiguanosina

**$\Phi$ 29DNAP:** DNA polimerasa del bacteriófago  $\Phi$ 29

**$\Phi$ 29TP:** TP del bacteriófago  $\Phi$ 29



**Fingers, palm y thumb:** subdominios que conforman el dominio de polimerización de las DNA y RNA polimerasas, llamados así por su similitud con los dedos, la palma y el pulgar de una mano derecha semiabierta

**Forward:** iniciador de la replicación directo

**Jelly-rolls:** estructura terciaria con forma de barril beta tipo remolino mostrada por la proteína mayoritaria de la cápsida de algunos virus como PRD1, Bam35 o adenovirus

**Jumping-back:** término empleado para describir el mecanismo de translocación hacia atrás del complejo de iniciación TP-(dNMP)<sub>3</sub> para recuperar la información genética de los tres primeros nucleótidos del extremo 3' de la cadena molde

**Pellet:** material precipitado

**Primer:** iniciador de la replicación. Puede ser un fragmento de DNA, RNA o un grupo hidroxilo de una proteína

**Reverse:** iniciador de la replicación reverso

**Sliding-back:** término empleado para describir el mecanismo de translocación hacia atrás del complejo TP-dNMP tras la reacción de iniciación para recuperar la información genética del primer nucleótido del extremo 3' de la cadena molde

**Spike:** estructuras protuberantes asociadas a algunas partículas virales. Su función es el reconocimiento y anclaje al hospedador. Traducido como *espículas* en esta tesis

**Stepwise sliding-back:** término empleado para describir el mecanismo de translocación hacia atrás en el que son necesarios varios *sliding-back* consecutivos para recuperar la información genética del extremo 3' de la cadena molde

**Upstream:** en esta Tesis, hacia el sentido 3' de la cadena molde

**Wild type:** tipo natural o silvestre, sin modificación genética





Los miembros de la familia *Tectiviridae* se clasifican en dos grupos: fagos líticos que infectan a bacterias Gram-negativas y fagos atemperados cuyo hospedador son bacterias Gram-positivas. Mientras que el primer grupo, y más concretamente PRD1, ha sido ampliamente estudiado, no se conoce en detalle la biología de los bacteriófagos pertenecientes al segundo grupo. Con el objetivo de ampliar el conocimiento científico de los tectivirus que infectan a Gram-positivas y establecer un nuevo modelo de estudio dentro de este grupo, el presente trabajo se focalizó en el estudio de la replicación del genoma y de la interacción entre todas las proteínas del bacteriófago Bam35.

La DNA polimerasa de Bam35 (B35DNAP), codificada por el marco de lectura abierto 5 (ORF5) de su genoma, pertenece a la familia B y presenta los subdominios TPR1 y TPR2, característicos de las DNA polimerasas (DNAPs) que inician la replicación del material genético utilizando una proteína terminal (TP) como *primer*. La caracterización bioquímica de esta enzima mostró una polimerización altamente procesiva asociada al desplazamiento de banda. Además, posee actividad exonucleasa 3'-5' que actúa de manera coordinada con la actividad polimerasa y que, junto a la especificidad de inserción del nucleótido dictado por la hebra molde, le confiere una fidelidad comparable con la de otras DNAPs replicativas, como las habitualmente empleadas para amplificar DNA.

A pesar de su alta fidelidad, la B35DNAP es también capaz de replicar un molde de DNA con sitios abásicos insertando, preferentemente, una adenina enfrente del daño y sin producir cambios en el marco de lectura. Además, la delección del subdominio TPR2, requerido para la procesividad, impide la extensión del *primer* más allá del daño.

Por otro lado, la caracterización funcional y bioquímica de la proteína codificada por el ORF4 del genoma de Bam35 demostró que dicha proteína es la TP viral, que actúa como *primer* en la replicación. La B35DNAP utiliza la tercera base del extremo 3' del DNA como molde para catalizar la formación de un enlace fosfoéster entre el nucleótido complementario y el grupo OH de la tirosina 194 de la TP. A continuación, el complejo de iniciación se puede translocar hacia atrás mediante un nuevo mecanismo de *jumping-back* de un nucleótido para recuperar la información genética del extremo.

Por último, el presente estudio muestra la información obtenida del análisis generalizado entre las interacciones de las proteínas codificadas por el genoma de Bam35. Los datos obtenidos nos permiten proponer a la proteína P17 como proteína minoritaria de la cápsida y a la P24 como proteína del pentón. Además, la proteína hidrolítica de la membrana interna P26 podría tener una función estructural clave, ya que interacciona tanto con otras proteínas de membrana como con proteínas del pentón y del vértice especial. Por último, en las demás interacciones observadas están implicadas 10 proteínas más de función aún desconocida.



The members of the *Tectiviridae* family can be subdivided into two groups: lytic bacteriophages that infect Gram-negative bacteria and temperate phages whose hosts are Gram-positive bacteria. While the first group, namely PRD1, has been deeply studied, the biology of the bacteriophages belong to the second group is not well known. Therefore, to extend the knowledge of the tectiviruses that infect Gram-positive bacteria and to establish a new working model for this group, the present work is focused on the study of the genome replication and the protein-protein interactions of the Bam35 bacteriophage.

The Bam35 DNA polymerase (B35DNAP), encoded by the open reading frame 5 (ORF5) of its genome, belongs to the B family and contains the TPR1 and TPR2 subdomains, which are the hallmark of DNA polymerases (DNAPs) that initiate the replication using a terminal protein (TP) as a primer. The biochemical characterization of this enzyme showed a highly processive polymerization coupled to strand displacement. In addition, the B35DNAP has a 3'-5' exonuclease activity that acts in a coordinated manner with the polymerization activity. This activity, together with the specificity of insertion of the nucleotide dictated by the template strand, provides to the B35DNAP a fidelity rate similar to other replicative DNAPs generally used to amplify DNA.

In spite of its high fidelity, the B35DNAP is able to replicate a template with abasic sites. To do this, this enzyme inserts, preferably, an adenine opposite the damage and resume replication without generating a frameshift. Moreover, deletion of the TPR2 subdomain, required for processivity, impairs primer extension beyond the abasic site.

On the other hand, the functional and biochemical characterization of the protein encoded by the ORF4 of the Bam35 genome, showed that this protein acts as a primer during the DNA replication. The B35DNAP uses the third nucleotide at the 3' end of the DNA as the template to catalyse the formation of the phosphoester bond of the complementary nucleotide to an OH group of the tyrosine 194 residue of the TP. Subsequently, the initiation complex can translocate backward through a single-nucleotide jumping-back mechanism to recover the genetic information of the genome ends.

Finally, the information obtained from the high-throughput analysis between the interactions of all the proteins codified by the Bam35 genome allows us to propose the P17 protein as the minor capsid protein of Bam35 and P24 as the penton protein. Besides, the P26 inner membrane hydrolytic protein could have an important structural function due to its interactions both with other inner membrane proteins and penton and special vertex proteins. Finally, in the rest of the observed interactions there are 10 more proteins with yet an unknown function.





# INTRODUCCIÓN

1



Todas las formas de vida, a excepción de algunos parásitos intracelulares, son hospedadores de virus y/o otros elementos genéticos “egoístas” como transposones o plásmidos (Koonin y Dolja, 2013). El grupo biológico de los virus, constituido por  $10^{31}$  partículas virales, es el más numeroso de la biosfera (Krupovic y Bamford, 2011) y desempeña un papel primordial en la regulación del número, la diversidad y la evolución de sus hospedadores, además de causar un impacto significativo en los ciclos biogeoquímicos (Suttle, 2007, Forterre y Prangishvili, 2009, Rohwer y Thurber, 2009).

Los virus presentan una gran diversidad en términos de mecanismos de replicación, de regulación de la expresión génica, de estructuras virales y de tamaño y estructura del genoma. De esta manera, se pueden distinguir genomas virales de ácido desoxirribonucleico (DNA) o ácido ribonucleico (RNA) tanto de cadena sencilla (ss) como doble (ds) y con un rango de tamaño que oscila desde menos de 2 Kb hasta más de 2 Mb (Krupovic y Bamford, 2010, Krupovic *et al.*, 2011, Koonin y Dolja, 2014).

Los virus que infectan arqueas, bacterias y eucariotas tienen, en general, características diferentes (Krupovic *et al.*, 2011, Koonin y Dolja, 2014). Entre los virus eucariotas, los más comunes son los virus de RNA, los elementos genéticos que se retrotranscriben (retrotransposones) y retrovirus que generalmente se integran en el genoma del hospedador (Kazazian, 2004, Goodier y Kazazian, 2008). Por otro lado, aunque los virus de ssDNA y dsDNA son menos comunes y presentan menos diversidad, los elementos “egoístas” de DNA están ampliamente distribuidos en este grupo celular (Koonin y Dolja, 2014). Por el contrario, la mayoría de los virus que infectan a arqueas y bacterias poseen genomas de dsDNA siendo, el segundo grupo más abundante, los virus de ssDNA. En arqueas y bacterias también se han observado algunos retrotransposones, pero no retrovirus, mientras que los virus de RNA son muy raros. Muchos de los virus de dsDNA y de ssDNA cuyo hospedador son bacterias alternan su modo de reproducción entre el ciclo lítico y lisogénico. En el ciclo lisogénico, el genoma viral permanece latente en el interior de su hospedador y se transmite de manera vertical en forma de provirus o profago.

Los virus que infectan a bacterias constituyen el grupo más abundante y diverso del mundo viral. Hace 100 años, en 1917, Félix d’Herelle utilizó por primera vez el término *bacteriófago* para describir un agente bacteriolítico invisible, que ya había sido mencionado en trabajos anteriores por el bacteriólogo británico Frederick Twort. Desde entonces, el estudio de los bacteriófagos, aunque con altibajos, se ha desarrollado permitiendo su aplicación biotecnológica con diversos objetivos, como son el diseño de fármacos, la síntesis de nuevas proteínas, la generación de nuevas vacunas de DNA y la detección de bacterias patógenas entre otros, así como su aplicación en el campo de la biomedicina a través de la denominada “terapia fágica” (revisado en Sharma *et al.* (2016)).

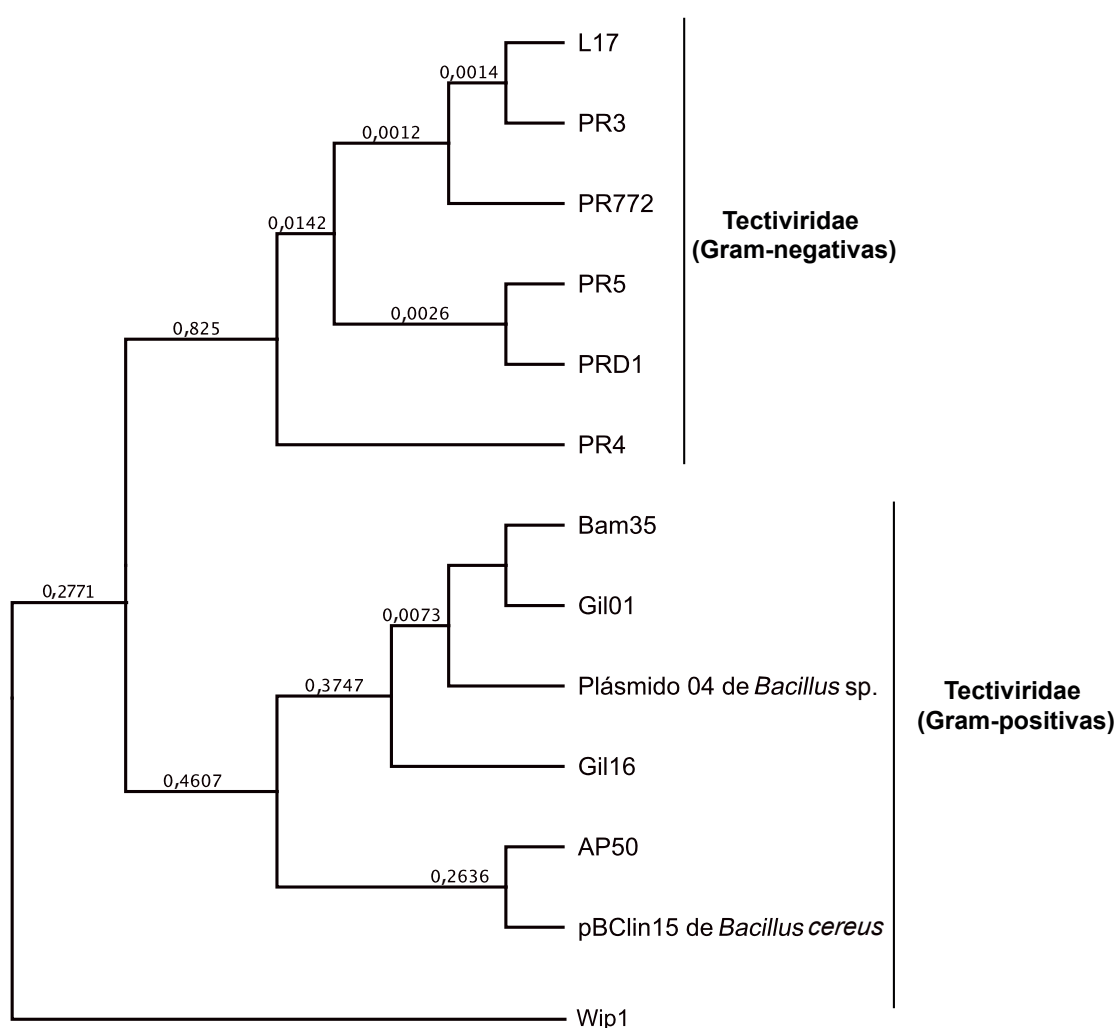
A pesar del amplio conocimiento de muchos modelos de bacteriófagos, como  $\Phi 29$  (Meijer *et al.*, 2001,

Serrano-Heras *et al.*, 2008, Ballesteros-Plaza *et al.*, 2013, Salas y de Vega, 2016) poco es conocido sobre la biología de otra gran cantidad de representantes, entre ellos algunos de los miembros de la familia *Tectiviridae*, objeto de estudio de esta tesis.



## 1. Familia *Tectiviridae*: características generales

La familia *Tectiviridae* está compuesta por virus bacterianos icosaédricos sin cola que presentan una membrana lipídica interna y un genoma de dsDNA lineal con repeticiones terminales invertidas (ITRs). Esta familia está formada por bacteriófagos que infectan a hospedadores muy diferentes pudiendo dividirla en dos grupos (**Figura 1**). En el primer grupo encontramos fagos líticos que infectan a enterobacterias Gram-negativas con plásmidos conjugativos como *Escherichia coli* (*E. coli*) o *Salmonella enterica* (Olsen *et*



**Figura 1. Análisis filogenético de las DNA polimerasas de los tectivirus.** El árbol se realizó mediante el alineamiento múltiple a través del programa ClustalW implementado en *Geneious* R8 (Kearse *et al.*, 2012) de las secuencias de DNA polimerasas de L17 (NCBI ID AAX45532.1), PR3 (AAX45563.1), PR772 (AAR99740.1), PR5 (AAX45625.1), PRD1 (NP\_040682.1), PR4 (AAX45594.1), Bam35 (NP\_943751.1), Gil01 (CAD59946.1), Plásmido 04 de *Bacillus* sp. (WP\_014905375.1), Gil16 (YP\_224103.1), AP50 (YP\_002302517.1), pBClin15 de *Bacillus cereus* (NP\_829893.1), Wip1 (YP\_008433325.1). A la derecha se indica al grupo de la familia *Tectiviridae* a la que pertenece cada representante.



al., 1974). Este grupo está compuesto por seis representantes que poseen un alta identidad en su secuencia de DNA: PR3, PR4, PR5, L17, PRR72 y PRD1, que es el estudiado con más detalle (Caldentey *et al.*, 1992, Caldentey *et al.*, 1993, Saren *et al.*, 2005, Karttunen *et al.*, 2015). El segundo grupo está formado por fagos atemperados que infectan bacterias Gram-positivas del grupo *Bacillus cereus* sensu lato y, aunque la secuencia de su genoma también está conservada, la identidad a nivel de nucleótido es más dispersa que en el grupo anterior (Verheust *et al.*, 2005, Kan *et al.*, 2013, Gillis y Mahillon, 2014a). Entre sus representantes se encuentran los bacteriófagos Bam35, Gil01 y Gil16, que infectan a *Bacillus thuringiensis* (Verheust *et al.*, 2005), AP50 y Wip1 cuyo hospedador es *Bacillus anthracis* (Nagy, 1974, Schuch *et al.*, 2010) y otros como Simila que infecta a *Bacillus cereus* (Gillis y Mahillon, 2014b). Además, estos virus están relacionados con el plásmido lineal de *B. cereus*, pBClin15, que podría ser un profago (Stromsten *et al.*, 2003a). A pesar de la diferencia de hospedadores y a la poca identidad de secuencia en su genoma, la mayor parte de los genomas de los miembros de esta familia muestran un tamaño de alrededor de 15 Kpb con una organización genética muy similar. Además, los bacteriófagos representativos de ambos grupos, PRD1 y Bam35, tienen una morfología viral casi indistinguible al microscopio electrónico. Estas coincidencias sugieren, fuertemente, un origen común para todos los tectivirus antes de la divergencia entre bacterias Gram-positivas y Gram-negativas (Ravanti *et al.*, 2003, Krupovic *et al.*, 2011).

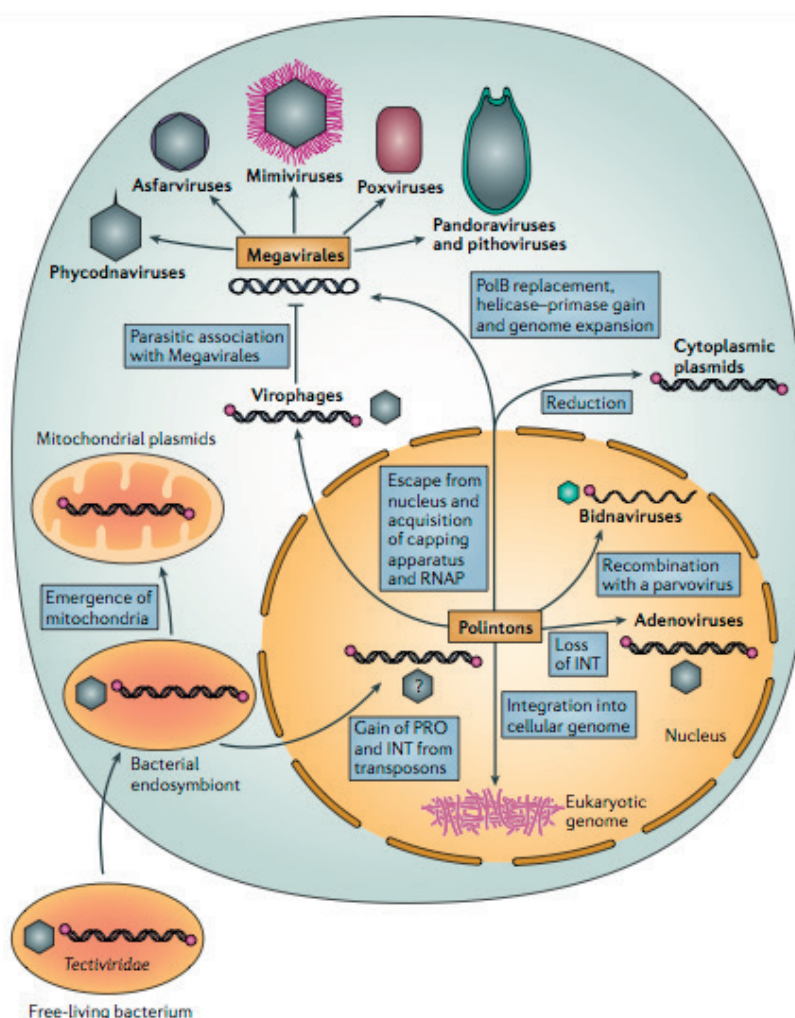


## 2. Relación filogenética de los tectivirus con otros grupos de virus

Aunque los tectivirus y los virus eucariotas de dsDNA infectan hospedadores de reinos muy diferentes, existen indicios que sugieren un origen evolutivo común. La primera evidencia de la relación filogenética entre estos virus fue descrita por Benson *et al.* (1999) al observar que la proteína mayoritaria de la cápsida de PRD1 (P3), al igual que su proteína equivalente en adenovirus, estaba formada por tres subunidades interconectadas constituidas por dos *jelly-rolls* con ocho láminas cada una. Además, ambos virus tienen una arquitectura general similar, con una cápsida que presenta la misma simetría y estructuras proteicas tipo fibras o espículas asociadas a los vértices. Otra de las similitudes es la presencia de una proteína terminal (TP) unida a los extremos 5' de sus genomas y de una DNA polimerasa (DNAP) perteneciente a la familia B (ver página 23).

Más recientemente, Krupovic y Koonin (2015) analizaron la relación filogenética entre unos transposones eucariotas de dsDNA de gran tamaño, llamados Polintones (o *Maverick*) y otros elementos genéticos móviles y virus de arqueas, bacterias y eucariotas. De este estudio se deduce una clara relación entre los polintones, los bacteriófagos pertenecientes a la familia *Tectiviridae* y adenovirus. Todos estos grupos tienen claros elementos comunes, como son los genes que codifican dos proteínas de la cápsida, la ATPasa de empaquetamiento de DNA y la DNAP de la familia B. Basándose en estas similitudes, los autores proponen

que los Polintones pudieron evolucionar desde un antecesor semejante a los tectivirus (**Figura 2**). Así, una bacteria ancestral habría sido infectada por un fago similar a los tectivirus. Posteriormente, una célula proto-eucariota adquiriría esta bacteria a través de un proceso de endosimbiosis. Una vez en el interior, la bacteria infectada podría haber evolucionado dando lugar a la mitocondria, explicando así la presencia de plásmidos lineales en su interior que codifican una DNAP de la familia B y tienen TPs en sus extremos. Por otro lado, el genoma del bacteriófago, podría haber sido importado al núcleo de la célula hospedadora, y eventualmente dar lugar a los Polintones y adenovirus. En línea con esta hipótesis, la presencia de señales de localización nuclear (NLS) funcionales en las TPs de diversos fagos, incluidos PRD1 y Bam35 pertenecientes a la familia *Tectiviridae* (Redrejo-Rodríguez *et al.*, 2012), podría ser una reminiscencia de la capacidad de estos virus para introducirse en el núcleo de la célula proto-eucariota.



**Figura 2. Esquema hipotético de la evolución de varios virus eucariotas y plásmidos.** Figura tomada de Krupovic y Koonin (2015). Según esta hipótesis, los Polintones evolucionarían desde un antecesor similar a los miembros de la familia *Tectiviridae* el cual entraría al hospedador proto-eucariota mediante una endosimbiosis bacteriana. A partir del Polinton, mediante diversos cambios y la adquisición de la capacidad de salir del núcleo emergerían los diversos virus eucariotas de DNA. Los hexágonos representan las cápsidas icosaédricas; en el caso del Polinton se marca con una interrogación porque se ha predicho, pero no comprobado, su existencia. Las cadenas dobles representan el dsDNA, mientras que la banda simple se refiere a ssDNA. INT es elemento retroviral de la familia integrasa RVE; PRO, proteasa del tipo de adenovirus.

Finalmente, un ancestro de los Polintones, habría adquirido diversos cambios entre ellos la capacidad de salir del núcleo, dando lugar a los virus grandes núcleo-citoplasmáticos eucariotas de DNA (NCLDV) y a virus satélites asociados a la familia de virus eucariotas gigantes *Mimiviridae* (La Scola *et al.*, 2008, Desnues *et al.*, 2012).



### 3. El bacteriófago Bam35 como modelo de estudio

El interés en el estudio de los tectivirus se ha incrementado recientemente debido a la estrecha especificidad de hospedador de algunos de ellos por patógenos humanos como *B. cereus* o *B. anthracis*, o por bacterias con capacidad insecticida como *B. thuringiensis* (Fischetti, 2011, Abulreesh *et al.*, 2012, Kan *et al.*, 2013), a su compleja diversidad y variabilidad entre sus miembros (Jalasvuori *et al.*, 2013, Gillis y Mahillon, 2014b) y a su posible relación con los virus eucariotas y otros elementos móviles descritos en el apartado anterior (Benson *et al.*, 1999, Krupovic y Koonin, 2015). Otro de los aspectos interesantes es la presencia de una membrana lipídica rodeada por una pared icosaédrica de proteínas, la cual no es habitual en los bacteriófagos pero está presente en algunos virus eucariotas como *Mimivirus*, poxvirus y el virus de la peste porcina africana (Cyrklaff *et al.*, 2005, Salas y Andrés, 2013, Klose y Rossmann, 2014), y es común en los virus de arqueas, como el virus *Sulfolobus* con forma de huso 1 (SSV1) del que recientemente se ha descrito un mecanismo de salida del hospedador muy similar al de virus eucariotas (Quemin *et al.*, 2016).

Por todo esto y porque, a diferencia de los tectivirus líticos que infectan a bacterias Gram-negativas, los tectivirus atemperados que infectan a Gram-positivas no han sido estudiados en detalle, nuestro laboratorio se interesó en establecer un nuevo modelo de estudio perteneciente a este grupo. Debido a la peligrosidad de las bacterias *B. anthracis* y *B. cereus*, los bacteriófagos AP50, Wip1 y  $\Phi$ NS11 fueron descartados como modelos. Entre los tres candidatos restantes (Bam35, Gil01 y Gil16), nos decidimos por el estudio de Bam35 debido a la disponibilidad de una mayor cantidad de datos previos, como estudios comparativos con PRD1 (Ravantti *et al.*, 2003), estudios estructurales (Laurinmaki *et al.*, 2005, Laurinavicius *et al.*, 2007) y del ciclo infectivo (Gaidelyte *et al.*, 2005, Daugelavicius *et al.*, 2007, Fornelos *et al.*, 2011).

#### 3.1. Morfología de la partícula viral

Las partículas virales de PRD1 y Bam35 (**Figura 3**) están compuestas por una cápsida externa proteica con simetría icosaédrica en cuyos vértices se encuentran ancladas unas espículas flexibles (*spikes*). A su vez, la cápsida encierra una membrana interna formada por proteínas codificadas por el virus y lípidos derivados de la membrana plasmática del hospedador celular (revisado en Oksanen y Bamford (2012)).

La cápsida de ambos virus está formada por 240 copias de una proteína trimérica, denominada proteína mayoritaria de la cápsida (MCP) (**Figura 3A**). Esta proteína es la P3 en el bacteriófago PRD1 mientras que

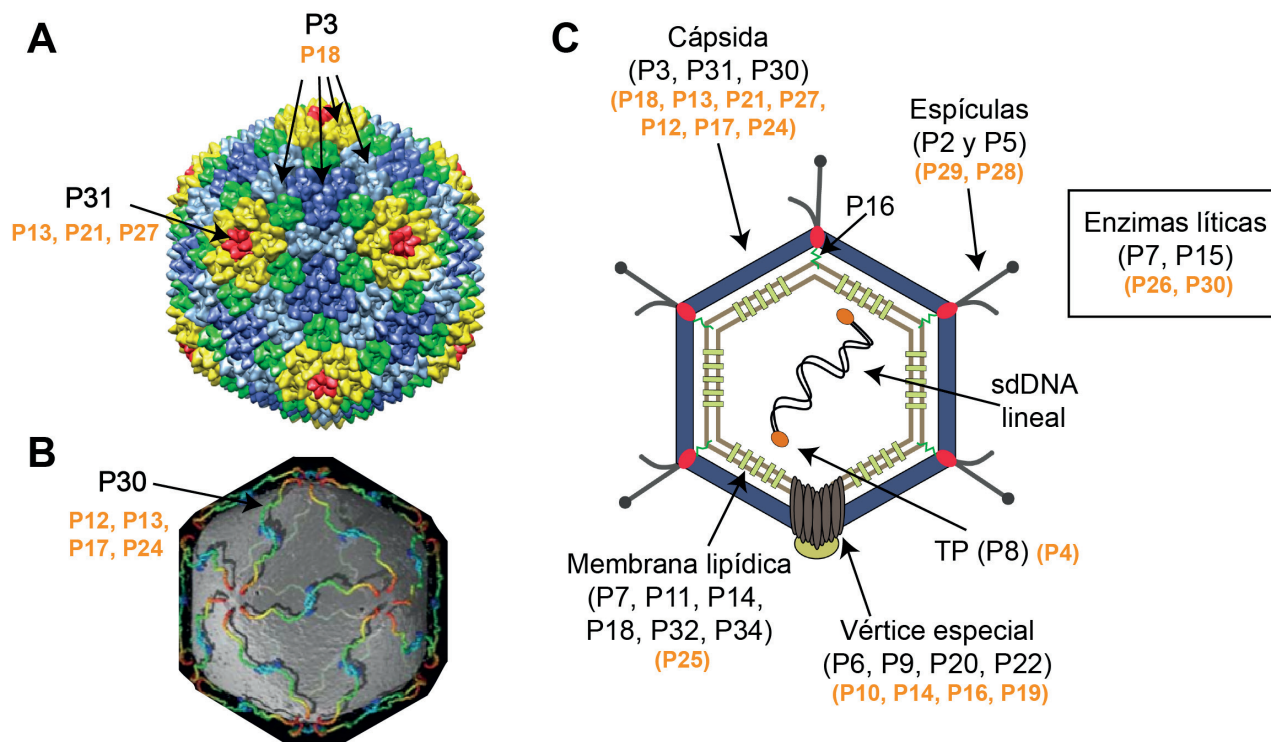
la P18 es su equivalente en Bam35 (43 y 39,7 KDa, respectivamente) (Ravanti *et al.*, 2003, Abrescia *et al.*, 2004, Laurinmaki *et al.*, 2005). El tamaño de la cápsida está determinado por dímeros de la proteína minoritaria de la cápsida que se localizan debajo de los capsómeros y conectan los vértices del virión (**Figura 3B**). En PRD1 la proteína encargada de esta función es la P30 (9,2 KDa) (Rydman *et al.*, 2001, Abrescia *et al.*, 2004). Sin embargo, hasta el momento, no se ha establecido su equivalente en Bam35, aunque por su tamaño se han sugerido como candidatas las proteínas P12, P13, P17 y P24 con 9,2, 12,2, 9,7 y 10,7 KDa, respectivamente (Laurinmaki *et al.*, 2005).

Las espículas se encargan del reconocimiento del receptor del hospedador y están ancladas a la cápsida a través de una estructura proteica denominada pentón (**Figura 3A y C**). El pentón de PRD1 está formado por la proteína P31 (14 KDa) (Rydman *et al.*, 1999) mientras que los candidatos propuestos para Bam35 son la P13 y la P21 por la similitud en el tamaño con su equivalente en PRD1 (12,2 y 16,3 KDa) y la P27 por la localización de su marco de lectura abierto (ORF) junto a las proteínas de la espícula (Laurinmaki *et al.*, 2005). Las espículas están formadas por dos proteínas (**Figura 3C**). Una de ellas es la proteína de unión al receptor, función llevada a cabo por la P2 de PRD1 y, probablemente, por la P29 en Bam35, a pesar de mostrar la mitad del peso molecular que su equivalente en PRD1 (32,7 KDa frente a 63,8 KDa) (Rydman *et al.*, 1999, Laurinmaki *et al.*, 2005). La otra proteína que compone la espícula de PRD1 (P5) presenta una estructura trimérica. Ravanti *et al.* (2003) proponen la P28 como su homólogo estructural en Bam35 al observar la presencia de una región de serinas y glicinas, similar al motivo tipo colágeno de la P5 de PRD1. Finalmente, el bacteriófago PRD1 muestra una pequeña proteína de membrana que une la espícula al complejo de la membrana (P16, 12,5 KDa) (Jaatinen *et al.*, 2004). Aunque los resultados obtenidos por Laurinmaki *et al.* (2005) sobre la estructura del virión de Bam35 indican que Bam35 tiene un homólogo a esta proteína, hasta el momento, no se ha sugerido ningún candidato posible.

Uno de los 12 vértices de la cápsida de PRD1 es diferente y contiene la maquinaria necesaria para el empaquetamiento del DNA en la partícula viral (**Figura 3C**). Este vértice se denominada “especial” y está formado por la ATPasa de empaquetamiento (P9), el denominado factor de empaquetamiento (P6) y por las proteínas P20 y P22 (25,7, 17,5, 4,6 y 5,5 KDa, respectivamente) (Stromsten *et al.*, 2003b, Stromsten *et al.*, 2005, Karhu *et al.*, 2007). Los posibles homólogos funcionales en Bam35 son las proteínas P14, P10, P16 y P19 respectivamente (24,6, 17, 5,3 y 8,3 KDa) (Ravanti *et al.*, 2003).

En el interior de la cápsida se encuentra la membrana lipídica asociada a proteínas virales (**Figura 3C**). En PRD1 la proteína mayoritaria de la membrana, P11, está implicada en la iniciación de la liberación del DNA mientras que la P18 es esencial para formar el tubo necesario para dicha liberación (22 y 9,6 KDa, respectivamente) (Grahm *et al.*, 2002, Mattila *et al.*, 2015). Además, existen otras proteínas de membrana

implicadas en la inyección del genoma como son la P14 y la P32 (15 y 5,4 KDa). En el caso de Bam35, el equivalente a la P11 de PRD1 posiblemente sea la P25 (22,4 KDa) (Ravanti *et al.*, 2003), mientras que para las proteínas P18, P14 y P32 no se ha propuesto ningún candidato.



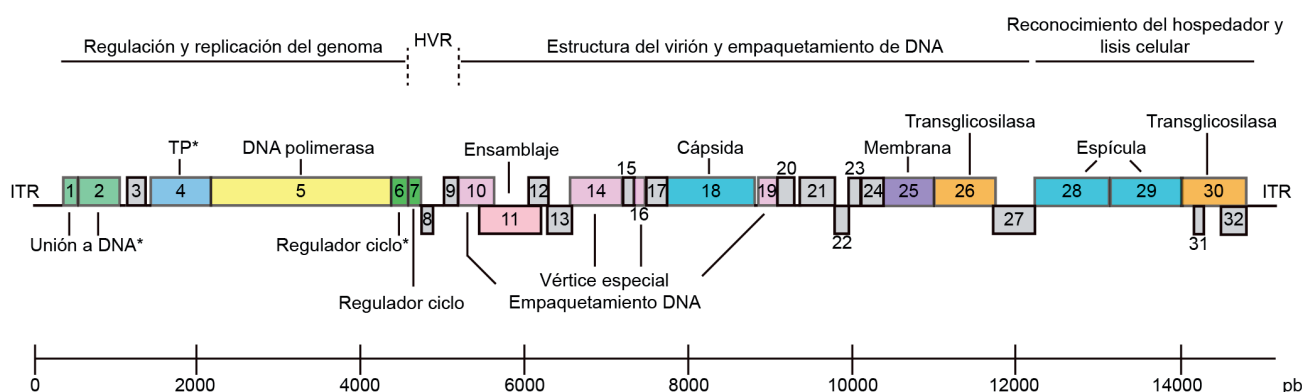
**Figura 3. Estructura del bacteriófago PRD1.** A) Estructura de la cápsida de PRD1 obtenida mediante cristalografía de rayos X (PDB 1W8X). La proteína P31 forma el pentón al que se ensambla la espícula (indicada en rojo). Las subunidades de la proteína P3 se organizan como un conjunto de cuatro trímeros posicionados en orientaciones diferentes las cuales están representadas por los colores verde, azul, azul claro y amarillo. B) Representación de la localización de la proteína P30 (Abrescia *et al.*, 2004). C) Representación esquemática de la partícula viral de PRD1. Las proteínas que forman la cápsida, la membrana interna, las espículas y el vértice especial así como la TP y su genoma están indicados. Las proteínas de PRD1 se indican en negro y sus equivalentes propuestos en Bam35 se indican en color naranja. Figura modificada de Oksanen y Bamford (2012).

Asimismo, la partícula viral de PRD1 contiene dos enzimas con capacidad lítica: la P7 y la P15 (27 y 17 KDa) (Mindich *et al.*, 1982, Pakula *et al.*, 1989, Caldentey *et al.*, 1994). La proteína P15 está asociada al virión, lo que es una característica extraña en sistemas líticos de virus-hospedador (Rydman y Bamford, 2000). Aunque se había sugerido su ubicación en el vértice especial, un estudio reciente no ha podido confirmarlo, por lo que su localización es aún un enigma (Mattila *et al.*, 2015). Como proteína equivalente a la P7 de PRD1 se ha propuesto la P26 en Bam35 (26 KDa). Esta proteína, al igual que la P30 de Bam35 (30,8 KDa), presenta actividad hidrolítica de peptidoglicano y está asociada a la partícula viral, del mismo modo que ocurre con las enzimas líticas de PRD1 (Gaidelyte *et al.*, 2006).



### 3.2. Genoma de Bam35

La bicapa lipídica encierra el genoma del bacteriófago, que en el caso de PRD1 es un dsDNA lineal de 14,9 Kpb con ITRs de 110 pb y una TP unida a ambos extremos 5' (Savilahti *et al.*, 1991, Caldentey *et al.*, 1992, Caldentey *et al.*, 1993). De esta manera su genoma se replica mediante un mecanismo iniciado por TP y mediante un sistema de deslizamiento hacia atrás (*stepwise silding-back*) evita la pérdida de la información genética de los extremos (ver página 27 y siguientes). Del mismo modo, el genoma de Bam35 es un dsDNA lineal de 14,9 Kpb con ITRs de 74 pb de longitud (Ravanti *et al.*, 2003) cuya imposibilidad para resolverse correctamente en un gel de agarosa sin tratar previamente con proteasas indica que está unido covalentemente a alguna proteína. Sin embargo, aunque el genoma de Bam35 muestra una clara similitud de secuencia con otros tectivirus de Gram-positivas, no ocurre lo mismo con respecto a los tectivirus de Gram-negativas u otros grupos de virus, dificultando la predicción o determinación de la función de las proteínas codificadas por sus ORFs (**Figura 4**). Este proceso es aún más complejo en el caso de las TPs puesto que sus secuencias están muy poco conservadas, incluso entre miembros de la misma familia. Generalmente, el gen que codifica la TP está situado al lado del que codifica la DNAP. Asimismo, las TPs suelen tener un alto punto isoeléctrico (pI) debido, sobre todo, a la acumulación de residuos con carga positiva en la mitad amino terminal (N-terminal) (pI de la TP de  $\Phi$ 29 es 8,6 y de PRD1 10,8). La TP de Bam35 y el gen que la codifica no han sido caracterizados hasta el momento, sin embargo, el ORF4 cumple con las dos características citadas previamente (pI de 10,6). Así, nuestro laboratorio y otros autores previamente, han sugerido que el ORF4 de Bam35 codifica la TP de este bacteriófago (Sozhamannan *et al.*, 2008, Redrejo-Rodríguez *et al.*, 2012, Jalasvuori *et al.*, 2013).



**Figura 4. Mapa genético de Bam35.** Los rectángulos corresponden a cada uno de los ORFs predichos y su número está indicado. Los rectángulos representados en la parte inferior indican que parte de ese ORF solapa con otro(s). Los diferentes colores representan el tipo de función de la proteína codificada por ese ORF. El color gris indica que la función es desconocida. El \* indica que las funciones de esas proteínas han sido predichas pero no demostradas experimentalmente. HVR, región altamente variable entre los tectivirus. La recta de debajo representa los pares de bases en el genoma de Bam35.

La presencia de ITRs, de una posible TP y de una DNAP de la familia B perteneciente al grupo de las que inician la replicación con TP (con subdominios *Terminal Protein Region 1* y *2*, ver página 23) sugieren fuertemente, que el genoma de Bam35 también se replica mediante un mecanismo iniciado por TP.



#### 4. Técnicas de estudio de la complejidad de los organismos y su aplicación en bacteriófagos

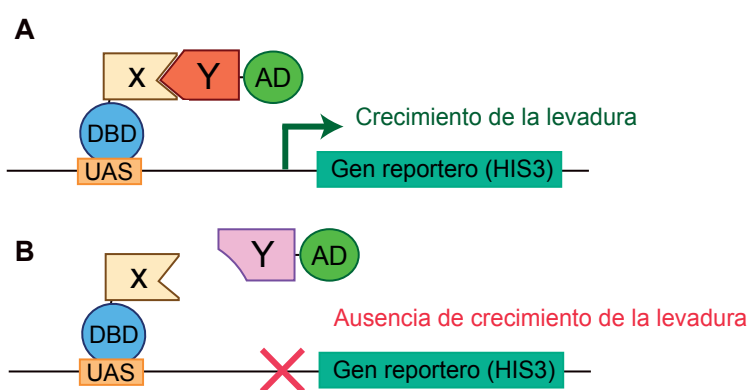
La existencia de gran diversidad genética entre los fagos dificulta la predicción de la función de sus proteínas por homología de secuencia. Sin embargo, otras características como las homologías estructurales o la sintenia pueden ser utilizada como base para elucidar su función.

Desde los inicios de la biología molecular de los fagos se han aplicado diferentes estrategias experimentales para determinar el papel biológico de las proteínas virales (revisado en Hauser *et al.* (2012)). Una de las primeras aproximaciones fue la denominada detección genética. Esta técnica está basada en el análisis de la pérdida de función en determinados mutantes y permitió, por ejemplo, determinar que la proteína J del fago  $\lambda$  se une a LamB de *E. coli* y que, por lo tanto, LamB es el receptor del fago  $\lambda$ . Posteriormente, con el diseño del primer microscopio electrónico por Ernst Ruska y Max Knoll, se facilitó el estudio de la relación física entre los componentes de los fagos. De esta manera, la combinación de la microscopía electrónica con análisis genéticos permite estudiar el cambio en la morfología de los bacteriófagos producidos por mutaciones en determinados genes. En la actualidad se está incrementando la aplicación de la microscopía crioelectrónica y la reconstrucción tridimensional que, junto a datos de alta resolución estructural permiten analizar la interacción entre proteínas localizadas en grandes complejos.

Otra de las técnicas que permite el estudio de la complejidad de los organismos consiste en la purificación del complejo proteico, objeto de estudio, para someterlo a análisis por espectrometría de masas. En el caso de los virus permite la identificación de las proteínas de la partícula viral. Otras técnicas implican la purificación de proteínas, lo que permite analizar sus interacciones y función mediante aproximaciones bioquímicas como, por ejemplo, *cross-linking*, co-inmunoprecipitaciones, ensayos de actividad, etc. Por otro lado, las proteínas individuales, complejos de proteínas y partículas virales purificadas pueden ser cristalizadas para someterlas a análisis por rayos X o a resonancia magnética nuclear (RMN). Estos métodos permiten determinar la estructura atómica de las proteínas y su interacción.

Otra técnica ampliamente aplicada en la detección de interacciones entre proteínas es el sistema de doble híbrido de levaduras, debido a su capacidad para revelar interacciones débiles y transitorias. Este sistema se ha utilizado para sugerir la función de proteínas codificadas por ORFs de función desconocida (ORFans) de diversos virus, como en el caso del bacteriófago  $\lambda$ , de Cp-1, del micobacteriófago Giles e incluso del virus de la Hepatitis E (Hauser *et al.*, 2011, Rajagopala *et al.*, 2011, Mehla *et al.*, 2015a, Osterman *et al.*, 2015). Además, se han aplicado en el análisis de las interacciones entre proteínas virales y las proteínas de sus hospedadores, utilizando librerías genómicas, como por ejemplo, entre las proteínas del bacteriófago  $\lambda$  y la librería de *E. coli*, o entre los bacteriófagos Dp-1 y Cp-1 con *Streptococcus pneumoniae* (Blasche *et al.*, 2013, Mariano *et al.*, 2016).

El sistema de doble híbrido en levaduras tradicional utiliza el factor de transcripción eucariótico del tipo Gal4p de *Saccharomyces cerevisiae* (Fields y Song, 1989). Este activador transcripcional tiene dos dominios funcionales: un dominio de unión al DNA en el N-terminal y un dominio activador en el carboxi-terminal (C-terminal). Cuando estos dos dominios están juntos se unen a la secuencia de activación *upstream* (UAS) de los genes del metabolismo de la galactosa en *S. cerevisiae*. Curiosamente, el factor Gal4p es funcional incluso si sus dominios no están covalentemente unidos. Esta característica es aprovechada en el sistema de doble híbrido en el cual se fusionan las proteínas objeto de estudio al dominio de unión al DNA de Gal4p (DBD), generando la proteína cebo, o al dominio de activación (AD), generando la proteína presa. Con el objetivo de incrementar la detección de interacciones positivas, los dominios de Gal4p se fusionan a las proteínas de estudio en distintas conformaciones (N- o C-terminal). Una de las construcciones que expresa una proteína cebo y otra de las que expresan una proteína presa se utilizaron para transformar la misma levadura, de manera que si las proteínas del fago interaccionan entre si el factor Gal4p será funcional y se unirá al UAS activando la expresión de un gen reportero que es específico y que proporciona un fenotipo seleccionable (ver esquema en **Figura 5**). En esta tesis el gen reportero utilizado fue el HIS3 que codifica una imidazol-glicerol fosfato deshidratasa que es crucial en la biosíntesis de la histidina permitiendo, con su expresión, el crecimiento de una levadura mutante para este gen en ausencia de este aminoácido. Una importante ventaja de la elección de este gen reportero es la posibilidad de disminución de falsos positivos mediante la adición del inhibidor competitivo del producto del gen HIS3, 3-amino-1,2,4-triazol (3AT) (Bruckner *et al.*, 2009).



**Figura 5. Esquema del sistema de doble híbrido.** Las proteínas cebo (X) se fusionan al dominio de unión al DNA (DBD) del factor de transcripción Gal4 que reconoce la secuencia de activación *upstream* (UAS) y la proteína presa (Y) a su dominio de activación (AD). Si ambas proteínas (X, Y) interaccionan, se activa la expresión del gen reportero, en este caso HIS3 permitiendo el crecimiento de la levadura en un medio de cultivo sin histidina (panel A). Si por el contrario las proteínas (X, Y) no interaccionan, el gen reportero no se expresa y como consecuencia no se observa crecimiento (panel B).



## 5. DNA polimerasas

La transmisión fiel de la información genética almacenada en el DNA es esencial para la supervivencia de todos los organismos. Para que esto sea posible, las DNA polimerasas (DNAPs) copian de manera precisa y fiel la cadena molde del genoma. Además, las DNAPs están implicadas en otros procesos destinados a mantener la integridad del material genético mediante la tolerancia al daño, la recombinación e incluso la mutagénesis.

A pesar de que las DNAPs se puedan clasificar en diferentes familias (ver apartado 5.1, página 23) todas ellas requieren de:

- Un extremo 3' iniciador, también denominado cebador o *primer*, ya sea DNA, RNA o una serina, treonina o tirosina de una proteína (Salas, 1991), que aporte un grupo OH al que unir los nucleótidos.
- Un molde que dicte el orden de incorporación de los nucleótidos durante la polimerización. No obstante, existen excepciones como la TdT, la Pol  $\mu$  o recientemente la Pol  $\theta$  (Bollum, 1960, Domínguez *et al.*, 2000, Kent *et al.*, 2016).
- Desoxinucleósidos trifosfato (dNTPs: dATP, dCTP, dGTP y dTTP) que al ser insertados liberan un pirofosfato quedando unidos al DNA naciente como desoxinucleósidos monofosfato (dNMP). Algunas DNAPs son capaces de utilizar también ribonucleósidos trifosfatos (rNTPs) con gran eficiencia, por ejemplo la TdT o la Pol  $\mu$  (Boule *et al.*, 2001, Ruiz *et al.*, 2003).
- Cofactores en forma de iones metálicos divalentes. Las DNAPs utilizan preferentemente el  $Mg^{2+}$ , aunque la catálisis también se produce en presencia de  $Mn^{2+}$  (Steitz *et al.*, 1994).

### 5.1. Familias de DNA polimerasas

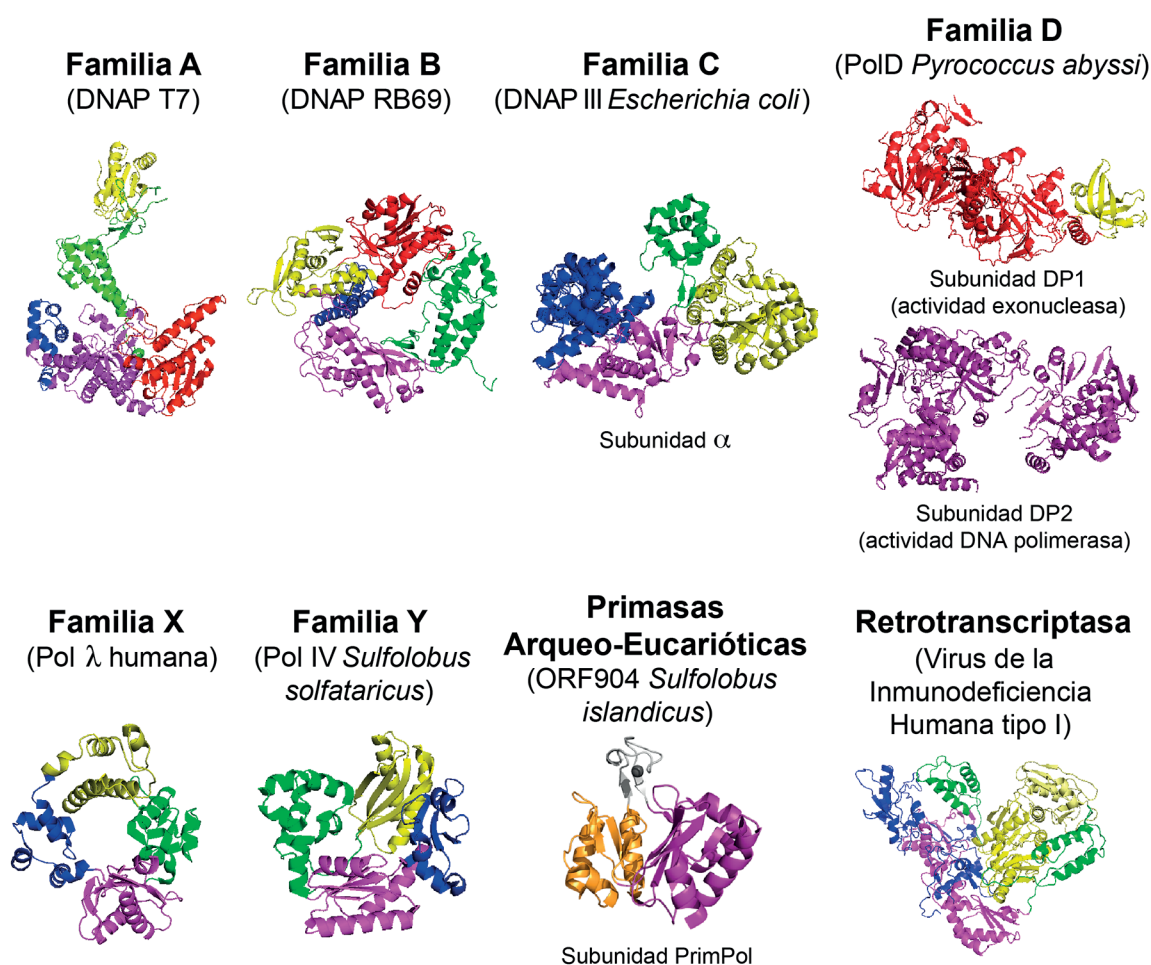
Las DNAPs se clasifican en diferentes familias dependiendo de: la estructura primaria de las subunidades catalíticas (Ito y Braithwaite, 1991, Braithwaite y Ito, 1993), de la sensibilidad a drogas e inhibidores y de la estructura tridimensional (Joyce y Steitz, 1994). Así, encontramos 8 familias diferentes (**Figura 6**):

- La familia A o DNAPs de tipo I: incluye DNAPs replicativas y de reparación. Algunos ejemplos de esta familia son las DNAPs de los bacteriófagos T3, T5 y T7. La DNAP mitocondrial eucariótica Pol  $\gamma$ , responsable de la replicación del DNA mitocondrial e implicada en procesos de reparación por escisión de bases (BER) (Bogenhagen *et al.*, 2001, Graziewicz *et al.*, 2006). Las Pol I de *E. coli*, de *Thermus aquaticus* (*Taq*) y de *Bacillus stearothermophilus*, además de la Pol  $\theta$  (Chan *et al.*, 2010, Yoon *et al.*, 2014) y la Pol  $\nu$  (Yamanaka *et al.*, 2010) humanas que están implicadas en procesos de reparación del DNA. Algunas de las DNAPs asociadas al procesamiento de fragmentos de Okazaki también pertenecen a esta familia.

- Familia B o DNAPs de tipo II: compuesta por DNAPs replicativas encargadas de replicar el genoma durante la división celular. Las DNAPs de esta familia poseen una potente actividad exonucleasa 3'-5' para realizar su función de la manera más fiel posible (Bernad *et al.*, 1989, Capson *et al.*, 1992). Algunos ejemplos de este grupo son las DNAPs celulares  $\alpha$ ,  $\delta$ ,  $\epsilon$  (implicadas en la replicación del DNA nuclear) y la  $\zeta$  (con capacidad de síntesis a través de lesiones), la DNAP bacteriana Pol II de *E. coli*, las de los fagos T4 y RB69 o las de los virus Vaccinia o Herpes; y las DNAPs que inician la replicación utilizando como *primer* una proteína, como la de  $\Phi$ 29 ( $\Phi$ 29DNAP) (Salas, 1991, Salas y de Vega, 2016), la de PRD1 (Caldentey *et al.*, 1992, Caldentey *et al.*, 1993), la de adenovirus (Hoeben y Uil, 2013) y, el modelo de estudio de esta tesis, la B35DNAP.
- Familia C: está compuesta por las DNAPs encargadas de replicar el genoma bacteriano. Algunos miembros de esta familia son la Pol III de *E. coli* y la de *Bacillus subtilis*.
- Familia D: incluye DNAPs del subdominio *Euryarchaeota* de arqueas. Probablemente, este tipo de DNAPs sean replicativas ya que la DNAP II de *Pyrococcus furiosus* mostró actividad polimerasa y exonucleasa 3'-5' (Uemori *et al.*, 1997).
- Familia X: compuesta por DNAPs implicadas en procesos de reparación del DNA (Uchiyama *et al.*, 2009) con una síntesis distributiva. Esta familia tiene representantes en mamíferos como la Pol  $\beta$  (Abbotts *et al.*, 1988),  $\lambda$  (García-Díaz *et al.*, 2000),  $\mu$  (Domínguez *et al.*, 2000) y TdT (Bollum, 1960); en arqueas y bacterias como la Pol X de *B. subtilis* (Baños *et al.*, 2008); y en virus como la de la peste porcina africana (Oliveros *et al.*, 1997).
- Familia Y: se inducen en presencia de agentes que causan daños en el DNA, poseen una síntesis de DNA distributiva y su eficiencia catalítica y fidelidad de inserción de nucleótido son muy bajos (Kunkel, 2004). Las DNAPs de esta familia están, en general, especializadas en sintetizar a través de un tipo de lesión particular (TLS). No muestran actividad exonucleasa 3'-5' (Zhou *et al.*, 2001) y al presentar el sitio activo más amplio, les permite aceptar diferentes tipos de lesiones e introducir nucleótidos con un par de bases incorrecto lo que se traduce en una baja fidelidad (Yang, 2005). Algunos ejemplos de esta familia son las DNAPs de bacterias, Pol IV y V de *E. coli*, de arqueas, Pol IV de *Sulfolobus solfataricus*, y las eucariotas Pol  $\iota$ ,  $\kappa$ ,  $\eta$  y Rev1 (Yang, 2003).
- Primasas Arqueo-Eucarióticas (AEP) (Iyer *et al.*, 2005): esta superfamilia incluye las primasas del grupo PrimPol, primasas que, además de sintetizar DNA de *nov*o, son capaces de extender *primers*. Esta característica nos permite clasificarlas también dentro de las DNAPs. Las PrimPol están implicadas en la iniciación de la replicación del material genético mitocondrial y nuclear, además de en la reparación del DNA gracias a su capacidad para sintetizar a través de lesiones (García-Gómez *et al.*, 2013, Martínez-Jiménez *et al.*, 2015). Su estructura terciaria no se asemeja al

resto de DNAPs (**Figura 6**).

- **Transcriptasas Reversas (RTs):** incluye las transcriptasas reversas de los retrovirus y las telomerasas eucarióticas. Utilizan un molde de RNA para sintetizar un DNA de doble cadena (Gotte *et al.*, 1999). En esta familia se encuentran, por ejemplo, la polimerasa monomérica del Virus de la Leucemia Murina de Moloney y la polimerasa dimérica del Virus de la Inmunodeficiencia Humana (Moelling *et al.*, 1971).



**Figura 6. Estructura terciaria de DNAPs pertenecientes a diferentes familias.** Se representan las DNAPs de los fagos T7 (PDB 1T7P), RB69 (1WAJ), la subunidad  $\alpha$  de la DNAP III de *E. coli* (2HQA), las dos subunidades de la PolD de *Pyrococcus abyssi* (5IHE y 5IJL), la DNAP  $\lambda$  humana (1XSN), la Pol IV de *S. solfataricus* (1JX4), la subunidad PrimPol de la proteína ORF904 codificada por el plásmido pRN1 (1RNI) y la retrotranscriptasa del Virus de la Inmunodeficiencia Humana tipo I (1DLO). Los subdominios *palm*, *thumb* y *fingers* se muestran en rosa, verde y azul, respectivamente. El dominio exonucleasa en rojo. El subdominio catalítico de la PolD y la proteína ORF904 se muestra en rosa. En la proteína ORF904 el subdominio de unión al DNA se marca en naranja y los dedos de zinc y el zinc en gris. Los subdominios amarillos son los específicos de cada DNAP: unión a tioredoxina en T7, N-terminal en RB69, el histidinol fosfatasa en Pol III, de unión al oligonucleótido o al oligosacárido en PolD, el de 8 KDa en Pol  $\lambda$ , el *little finger* en Pol IV y el de conexión y RNasa H en la retrotranscriptasa. Las estructuras se generaron con el programa *Pymol* (<http://pymol.org>).

## 5.2. Estructura de las DNA polimerasas

La mayoría de las DNAPs presentan su dominio de polimerización con una disposición estructural que se asemeja a una mano derecha semiabierta y que se divide en tres subdominios (**Figura 6**):

- *Palm*, es el más conservado. En él se encuentra el centro activo de polimerización con los aminoácidos que coordinan los iones metálicos catalíticos necesarios para la adición de nucleótidos (Steitz *et al.*, 1994). También contiene los residuos implicados en la unión del extremo 3' del *primer* y del fosfato  $\alpha$  del dNTP entrante (Joyce y Steitz, 1995).
- *Fingers*, este subdominio interacciona con la cadena molde del DNA para posicionarla en el centro activo (Kohlstaedt *et al.*, 1992, Jacobo-Molina *et al.*, 1993, Wang *et al.*, 1997). Asimismo, contiene residuos básicos que se encargan de interaccionar electrostáticamente con los fosfatos del dNTP entrante lo que permite su correcta orientación y estabilización.
- *Thumb* contiene residuos que interaccionan con el DNA para conferir estabilidad en la unión al extremo del *primer* (Kamtekar *et al.*, 2004). En las DNAPs con actividad exonucleasa es el responsable de la coordinación entre la actividad de polimerización y la de exonucleólisis (Beese *et al.*, 1993, Minnick *et al.*, 1996).

Específicamente, las DNAPs que tienen actividad de corrección de errores 3'-5' poseen un dominio adicional N-terminal que contiene la actividad exonucleasa.

Finalmente, las DNAPs que inician la replicación de su genoma utilizando una proteína como *primer* (TP), poseen dos dominios adicionales denominados *Terminal Protein Region 1* y *2* (TPR1 y TPR2) que están localizados en su dominio de polimerización (Blasco *et al.*, 1990, Dufour *et al.*, 2000).

- El subdominio TPR1 se localiza entre el *palm* y el *fingers*. La estructura cristalográfica de la  $\Phi$ 29DNAP y la TP de  $\Phi$ 29 muestra que este subdominio interacciona con la TP y con el dsDNA (Kamtekar *et al.*, 2004, Berman *et al.*, 2007).
- El subdominio TPR2 se localiza enfrente del *thumb* formando una estructura similar a un arco. Junto al *palm* y *fingers* rodean la hebra molde formando un estrecho túnel cuyas dimensiones no permiten la entrada de dsDNA forzando su apertura para permitir que la hebra molde llegue al centro de polimerización. Esto favorece el desplazamiento de banda por la DNAP, por ejemplo en la  $\Phi$ 29DNAP, sin necesidad de la presencia de helicasas (Kamtekar *et al.*, 2004, Rodríguez *et al.*, 2005).

Los subdominios *thumb*, *palm*, TPR1 y TPR2 interaccionan con el dúplex de DNA naciente en el sitio de polimerización (Berman *et al.*, 2007) proporcionando a la DNAP la máxima estabilidad de unión al DNA, la cual es requerida para la procesividad intrínseca.



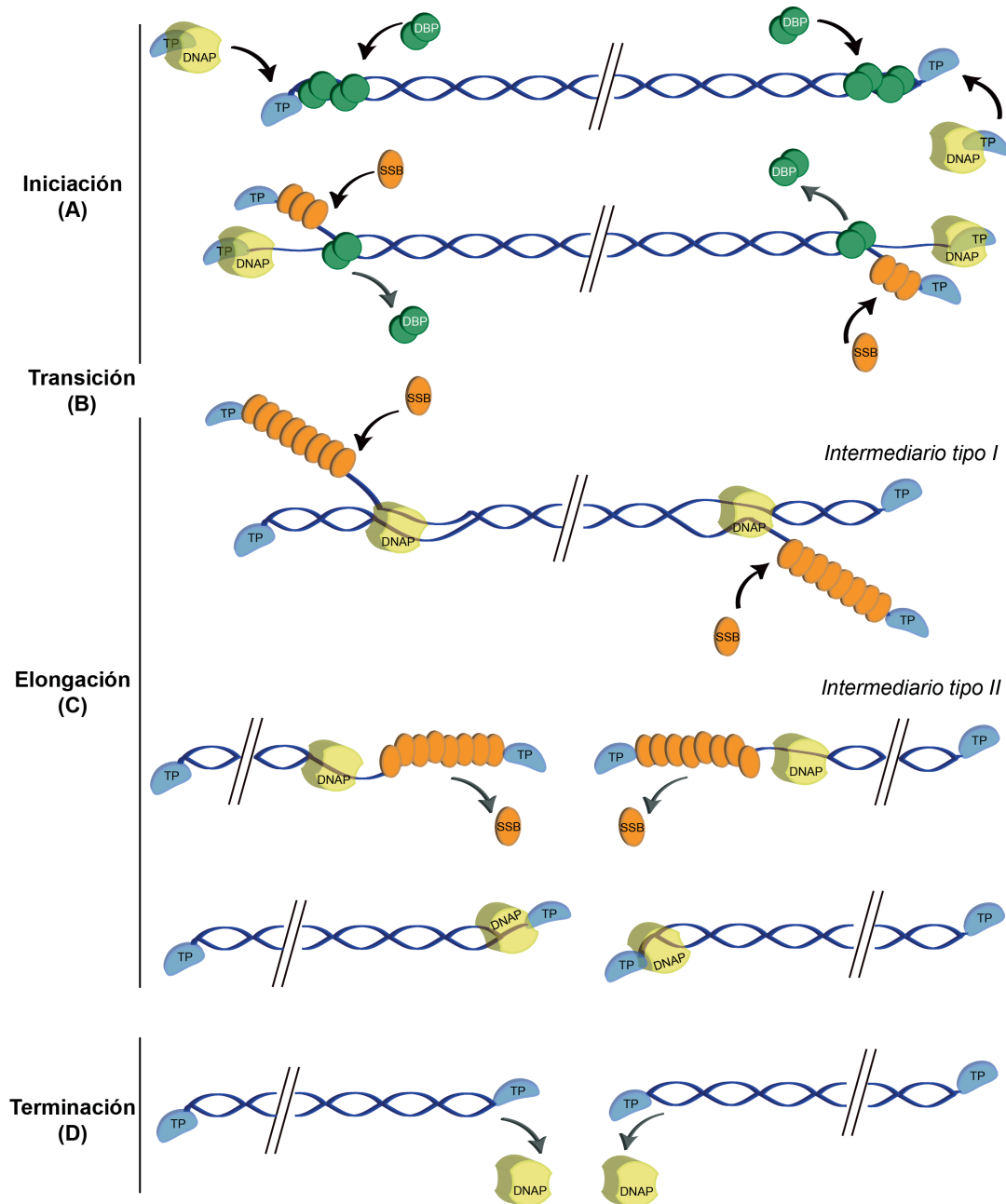
## 6. Replicación iniciada por proteína terminal

Como ya se mencionó previamente, algunas DNAPs de la familia B son capaces de iniciar la replicación del genoma utilizando un grupo OH perteneciente a una serina, treonina o tirosina de una proteína, denominada TP (Salas, 1991). Mediante homología de secuencia, se determinó que el ORF5 del genoma de Bam35 codifica una DNAP perteneciente a esta familia y muestra los dos subdominios característicos de ese mecanismo de replicación (TPR1 y TPR2). Por otro lado, su dsDNA contiene ITRs y una proteína unida covalentemente (Ravanti *et al.*, 2003). Estas características sugieren que la B35DNAP podría utilizar una TP para iniciar la replicación del genoma, al igual que la DNAP del bacteriófago PRD1, perteneciente a su misma familia (Caldentey *et al.*, 1992).

El modelo más estudiado con este tipo de replicación es el bacteriófago  $\Phi 29$  (Salas y de Vega, 2016). No obstante, este mecanismo está ampliamente distribuido, encontrándose, además de en bacteriófagos, en virus animales como adenovirus o el virus humano de la hepatitis B, en plásmidos mitocondriales y citoplasmáticos de levaduras y hongos filamentosos, así como en cromosomas lineales y plásmidos de *Streptomyces* (Salas, 1999). Asimismo, debido a la presencia de los subdominios TPR1 y TPR2 en las DNAPs, se ha sugerido este tipo de replicación en virus que infectan a arqueas (Bamford *et al.*, 2005, Bath *et al.*, 2006, Peng *et al.*, 2007), además de en elementos de transposición de eucariotas y algunos virofagos (Kapitonov y Jurka, 2006, Fischer y Suttle, 2011).

En el caso de  $\Phi 29$ , la replicación se inicia de manera no simultánea desde ambos extremos del genoma y comienza con la formación de un heterodímero constituido por la  $\Phi 29$ DNAP y la TP iniciadora ( $\Phi 29$ DNAP/TP). Este complejo reconoce los extremos del DNA unido a la TP (TP parental) denominados orígenes de replicación (**Figura 7A**). Al mismo tiempo, las proteínas de unión al dsDNA (DBP), codificadas por el genoma del fago, forman un complejo nucleoproteico en los extremos del DNA. Esta unión facilita la apertura de la hélice y, en presencia del nucleótido iniciador, en este caso dATP, complementario a la segunda base de la hebra molde (3'-TTT), estimula la formación por la  $\Phi 29$ DNAP de un enlace fosfoéster entre el nucleótido y el grupo OH de la serina 232 de la TP iniciadora. Posteriormente, se produce la translocación hacia atrás del complejo iniciador (TP-dAMP) para recuperar la información genética del primer nucleótido del genoma (3'-T), mediante un mecanismo denominado *sliding-back* (ver página 29). A continuación, la  $\Phi 29$ DNAP, aún unida a la TP, cataliza la polimerización de diez nucleótidos, en una etapa denominada transición (**Figura 7B**). Seguidamente, la  $\Phi 29$ DNAP se disocia de la TP y continúa con la elongación de manera procesiva acoplada al desplazamiento de la cadena complementaria al molde sin necesidad de helicasas ni factores de procesividad (**Figura 7C**). Durante esta etapa se generan los intermediarios de replicación tipo I consistentes en moléculas de dsDNA de tamaño completo con una o más colas de

ssDNA cubiertas por las proteínas de unión al ssDNA (SSB) codificadas por el genoma del fago. Cuando las horquillas de replicación que se iniciaron de manera asimétrica desde los extremos opuestos del genoma convergen, los intermediarios de replicación tipo I se convierten en intermediarios de tipo II en los que la zona por donde comenzó la replicación es de dsDNA mientras que su opuesta es de ssDNA. La replicación completa de este intermediario conlleva la disociación de las proteínas SSB y da lugar a dos moléculas de dsDNA. La disociación final de la  $\Phi 29$ DNAP de cada molécula de DNA permite la formación de nuevos heterodímeros con las TPs iniciadoras libres para dar lugar a una nueva ronda de replicación (**Figura 7D**).



**Figura 7. Esquema del mecanismo de replicación del genoma de  $\Phi 29$ .** La iniciación se produce de manera no simultánea desde ambos extremos. El nombre de cada una de las proteínas que participan en las diferentes etapas y de los intermediarios de tipo I y II están indicados. La etapa de transición no está representada. Figura modificada de De Vega y Salas (2011).



En el caso de  $\Phi 29$  o PRD1 la replicación procesiva del genoma *in vitro* se puede llevar a cabo con la DNAP y la TP sin necesidad de proteínas adicionales, aunque su presencia incrementa la replicación y es esencial *in vivo* (Pakula *et al.*, 1990, Salas, 1991, Pakula *et al.*, 1993). Sin embargo, en otros modelos, como adenovirus o Nf, la presencia de proteínas adicionales es también esencial para la replicación *in vitro* (Dekker *et al.*, 1997, Longás *et al.*, 2006).

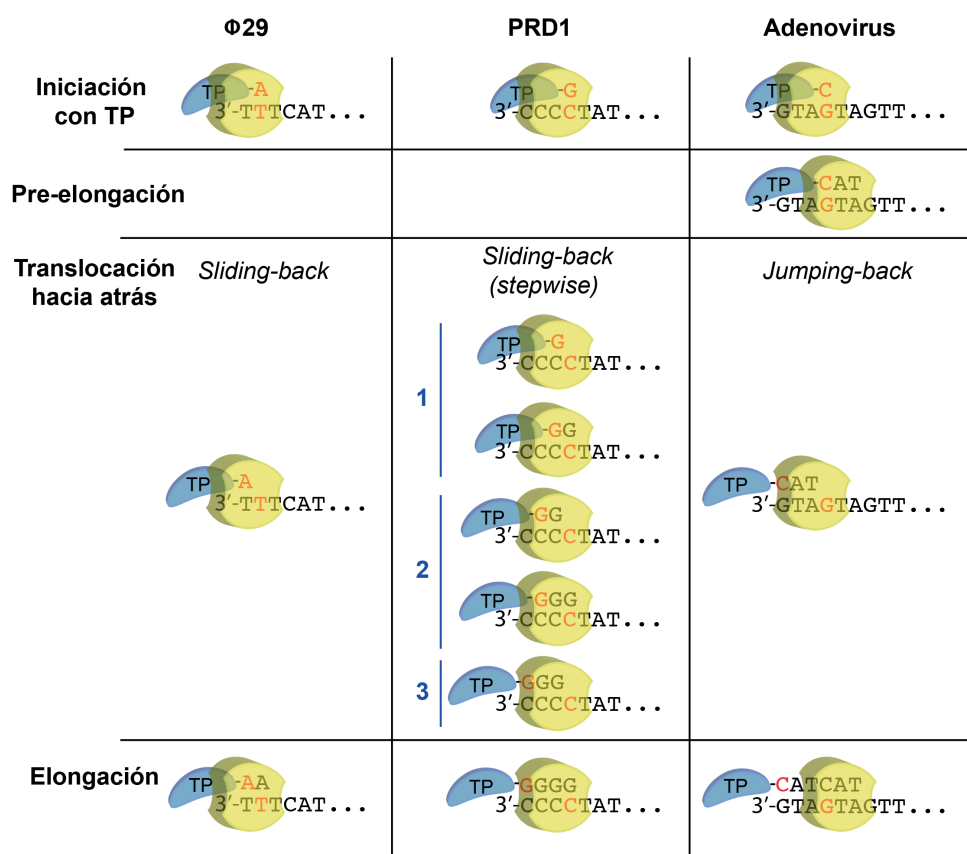
### 6.1. Mecanismos de recuperación de la información genética de los extremos 3' del genoma en la replicación iniciada por TP

La replicación iniciada por TP suele estar dirigida por la segunda, tercera o cuarta base de la cadena molde del DNA. El material genético de este tipo de replicación presenta repeticiones en los extremos. De esta manera, tras el inicio de la replicación en una posición interna del DNA, el complejo formado por la TP unida a dNMP(s) es capaz de translocarse hacia atrás e hibridar con la secuencia del extremo 3' del molde. Además, gracias a la necesidad de la complementariedad entre el/los dNMP(s) unidos a la TP y la secuencia del extremo 3' del molde para una translocación correcta, se previenen las mutaciones en los extremos del genoma al dificultar que un producto erróneo de iniciación sea utilizado por la DNAP para continuar con la replicación (Esteban *et al.*, 1993, Salas y de Vega, 2016). Esta característica es de gran importancia puesto que la actividad exonucleasa de las DNAPs no es capaz de actuar sobre sustratos tan pequeños.

Hasta el momento se han descrito tres mecanismos de recuperación de la información genética según la secuencia de los extremos del genoma:

- **Sliding-back (Figura 8, izquierda):** En este caso la DNAP cataliza la formación de un enlace fosfoéster entre el nucleótido complementario a la segunda base del extremo 3' de la hebra molde del DNA y un grupo OH de una serina, treonina o tirosina de la TP. Para que este mecanismo de recuperación sea posible, la secuencia del extremo de la hebra molde tiene que presentar dos pares de bases iguales. De esta manera, tras la iniciación, el producto TP-dNMP se translocaría hacia atrás hibridando con la primera base del DNA para recuperar la información genética del extremo 3'. Finalmente, la DNAP cataliza la formación de un enlace fosfodiéster entre el nucleótido complementario a la segunda base de nuevo y el grupo 3'OH del dNMP unido a la TP. Este mecanismo es el utilizado por  $\Phi 29$  y GA-1 cuya secuencia 3' del genoma es 3'-TTT (Méndez *et al.*, 1992, Illana *et al.*, 1996).
- **Stepwise sliding-back (Figura 8, centro):** este mecanismo consiste en varios *sliding-back* consecutivos, ya que la iniciación se produce utilizando como molde la tercera o cuarta base del extremo 3' del genoma. De esta manera, después de la catálisis por la DNAP del enlace fosfoéster entre la TP y el nucleótido, el complejo TP-dNMP se transloca hacia atrás una

posición. Posteriormente, la DNAP copia de nuevo el primer nucleótido molde generando el enlace fosfodiéster con el grupo 3'OH del dNMP unido a la TP. A continuación, se produce otro *sliding-back* de una posición, y la DNAP vuelve a copiar el mismo nucleótido, y así hasta que se copia todo el extremo del genoma y la DNAP puede continuar replicando el molde. Este mecanismo es utilizado por PRD1 que inicia su replicación en la cuarta posición del extremo 3' del genoma (3'-CCCC) necesitando 3 pasos de *sliding-back* para recuperar la información (Caldentey *et al.*, 1993).



**Figura 8. Mecanismos de recuperación de la información genética de los extremos 3' del genoma en la replicación iniciada por proteína terminal.** El esquema muestra los mecanismos de *sliding-back* (Φ29), *stepwise sliding-back* (PRD1) y *jumping-back* (adenovirus). La TP se representa en color azul y la DNAP en amarillo. Los nucleótidos en rojo marcan la base utilizada como molde en la iniciación y el primer nucleótido insertado. Figura modificada de De Vega y Salas (2011).

- *Jumping-back* (**Figura 8**, derecha): se produce cuando la secuencia de los extremos 3' tienen una reiteración más compleja, por ejemplo, 3'-GTAGTA de adenovirus. En este caso, la DNAP comienza la replicación enfrente de la cuarta base del extremo 3', catalizando la formación del enlace fosfoéster entre el nucleótido complementario y la TP (generando TP-dCMP, TP-C). A continuación, la DNAP replica hasta la sexta base formando el complejo TP-CAT que, posteriormente, salta hacia atrás para recuperar la información genética del extremo 3' mediante el apareamiento con la primera, segunda y tercera base (3'-GTA). Finalmente, la DNAP continúa con la replicación del genoma (King y van der Vliet, 1994).





## 7. Fidelidad en la replicación del DNA

Un apropiado nivel de fidelidad durante la replicación del DNA asegura la transferencia de información genética de una generación a la siguiente. Es esencial que la duplicación semiconservativa del DNA de lugar a un producto final casi perfecto ya que, de otro modo, algunos genes importantes podrían tener mutaciones que darían lugar a enfermedades o a la muerte celular. Por ello, las DNAPs muestran distintos mecanismos para asegurar una replicación fiel (revisado en Ganai y Johansson (2016)):

- La selección del nucleótido correcto: es el mayor determinante en la fidelidad. Las DNAPs forman un complejo con el DNA seguido de la unión de un nucleótido y dos iones metálicos en el centro activo de polimerización de la enzima (Freudenthal *et al.*, 2012). Esta unión depende de las interacciones entre el nucleótido y la base presente en el molde, de los residuos del centro activo, del subdominio *fingers*, de los iones metálicos y del grupo 3'OH del *primer*. Estas interacciones no se completan hasta que hay un cambio conformacional en la DNAP, provocado por un mecanismo de ajuste inducido por el sustrato. Si la DNAP trata de incorporar un nucleótido incorrecto, la pérdida de la geometría correcta en el sitio activo imposibilitaría la formación del enlace fosfodiéster y el nucleótido sería descartado.
- Corrección de pruebas: este mecanismo se divide en dos pasos; primero, la DNAP reconoce un nucleótido mal incorporado y, segundo, el nucleótido es retirado por la actividad intrínseca exonucleasa 3'-5' presente en las DNAPs replicativas. La fidelidad de las DNAPs replicativas depende del balance entre sus actividades polimerasa y exonucleasa y de su capacidad para elongar un nucleótido no apareado en el extremo 3' de la hebra naciente. Así, la inserción del nucleótido correcto favorece la replicación mientras que ante un nucleótido incorrecto la polimerización se bloquea y la cadena naciente se transfiere al sitio activo de exonucleasa. En el caso de no poseer actividad exonucleasa o esta no ser capaz de eliminar el nucleótido desapareado, la DNAP se disocia y otras proteínas se encargarían de retirar el nucleótido erróneo. Para detectar la inserción de nucleótidos erróneos, las DNAPs de la familia B tienen motivos conservados que contactan con los últimos pares de base generados, como el motivo KxY del *palm* y otros residuos del *thumb* (Blasco *et al.*, 1995, Pérez-Arnaiz *et al.*, 2006).
- Reparación de errores de apareamiento (*mismatch repair*, MMR): este mecanismo se utiliza para reparar los errores que escapan a la actividad exonucleasa 3'-5' de las DNAPs (Kunkel y Erie, 2015). Este sistema de reparación se lleva a cabo por proteínas específicas que reconocen errores de apareamiento, inserciones y deleciones durante la replicación.

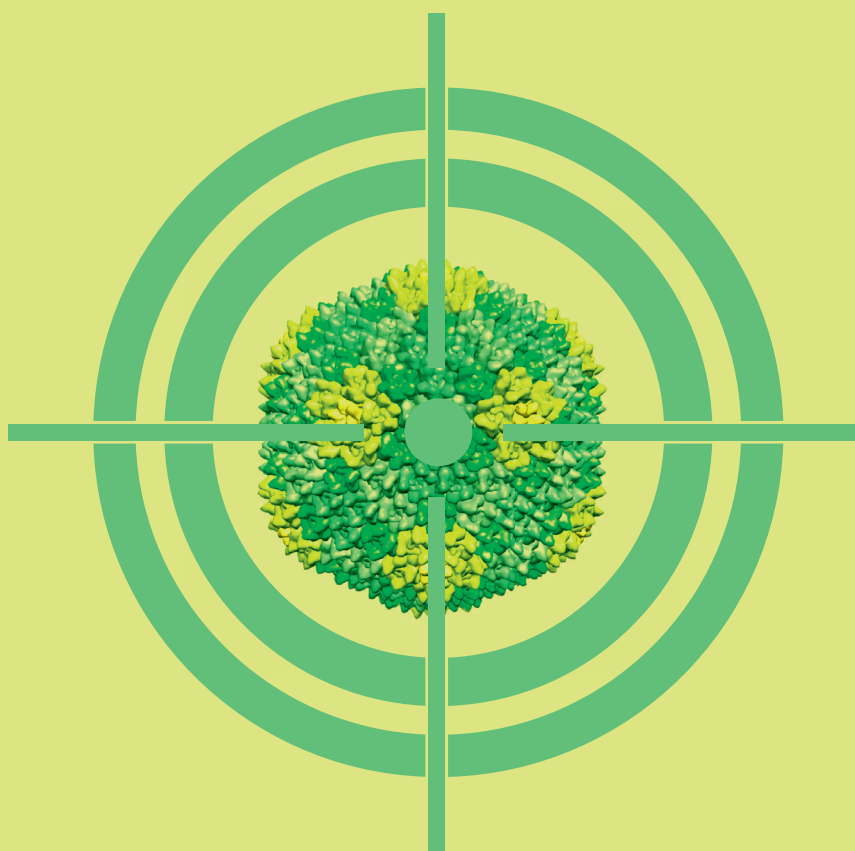
### 7.1. Síntesis de DNA a través de lesiones

A pesar de la presencia de esos mecanismos de fidelidad en las DNAPs, la estabilidad del genoma puede verse comprometida por la presencia de daños producidos por agentes mutagénicos ambientales mediante reacciones de alquilación, oxidación, desaminación, hidrólisis, foto-adición, etc. (revisado en Liu *et al.* (2016)). Estas lesiones en el DNA pueden reducir la fidelidad de polimerización, causar mutaciones que cambien el marco de lectura o incluso bloquear la replicación. Las lesiones que no bloquean la replicación pueden ser superadas por algunas DNAPs replicativas. Un ejemplo es la lesión más común producida por estrés oxidativo, 7,8-dihidro-8-oxo-2'-oxodeoxiguanosina (8-oxoG), que dependiendo de la DNAP, del contexto de secuencia y de la disponibilidad de dNTPs puede dar lugar a un producto sin o con mutaciones (normalmente dAMP en lugar de dCMP) (De Vega y Salas, 2007, McCulloch *et al.*, 2009). Por el contrario, las DNAPs replicativas generalmente no sintetizan a través de lesiones bloqueantes debido al "estrecho ajuste" del DNA y el dNTP entrante en el sitio activo de la proteína, que está muy optimizado para la formación de un apareamiento Watson-Crick (Johansson y Dixon, 2013, O'Donnell *et al.*, 2013). Por ello, al alcanzar este tipo de lesiones la horquilla de replicación se para y, para poder superarlo, la DNAP replicativa es sustituida por una DNAP de síntesis a través de lesiones (TLS), normalmente perteneciente a la familia Y. Estas DNAPs muestran un centro activo sin solventes expuestos lo que les permite encajar DNAs con características aberrantes llevando a cabo, como consecuencia, una polimerización de baja fidelidad (McCulloch y Kunkel, 2008, Yang, 2014). Una vez sintetizado el daño, la DNAP de TLS es sustituida por la DNAP replicativa para continuar la polimerización. Este tipo de síntesis puede dar lugar a tres tipos de procesos: genoma sin mutación, con mutación o muerte celular (Liu *et al.*, 2016).

La lesión bloqueante más común del DNA son los sitios abásicos o apurínicos/apirimídicos que se producen cuando el enlace N-glicosídico entre el azúcar y la nucleobase se rompe de manera espontánea o como intermediario de la reparación del DNA por escisión de bases (Lindahl, 1993, Nakamura y Swenberg, 1999). El estudio de la capacidad de síntesis a través de estos daños por las DNAPs replicativas ha demostrado que dicha síntesis es contrarrestada por su actividad exonucleasa 3'-5'. De esta manera, simplemente la eliminación de esta actividad en algunas DNAPs de la familia B como en la del bacteriófago T4 o del virus del herpes simple, les proporciona la capacidad de replicar un DNA con sitios abásicos (Tanguy Le Gac *et al.*, 2004, Zhu *et al.*, 2008). Por otro lado, la inserción de variaciones en los residuos del sitio activo de las DNAPs replicativas también permite la síntesis a través de sitios abásicos al afectar a la selección del nucleótido entrante. Así, por ejemplo, la modificación única de la leucina 868 del subdominio *palm* de la Pol  $\alpha$  le proporciona dicha capacidad (Niimi *et al.*, 2004). Sin embargo, de manera excepcional, algunas DNAPs replicativas, como la procedente de un polintón de *Entamoeba histolytica* tienen cierta capacidad intrínseca de sintetizar a través de sitios abásicos *in vitro*, siendo esta capacidad dependiente de su procesividad

(Pastor-Palacios *et al.*, 2012).





## OBJETIVOS

2



Los bacteriófagos que infectan a bacterias Gram-positivas y que pertenecen a la familia *Tectiviridae* no han sido profundamente estudiados a pesar del reciente incremento en el interés por esta familia. Por ello, el objetivo principal de esta tesis fue ampliar el conocimiento del sistema replicativo de este grupo de bacteriófagos utilizando como modelo Bam35, y más concretamente centrándonos en:

1. La caracterización bioquímica de la B35DNAP mediante el estudio de sus actividades de polimerización y exonucleolisis, así como de la capacidad de desplazamiento de la cadena de DNA que no actúa como molde y de la fidelidad durante la replicación.
2. La caracterización funcional y bioquímica de la proteína codificada por el ORF4 del genoma del bacteriófago Bam35.
3. El estudio de las primeras etapas de la replicación de DNA iniciada por TP. Determinación de la posición del nucleótido de la cadena molde del DNA utilizado por la B35DNAP para la iniciación, y elucidación del mecanismo de recuperación de la información genética de los extremos 3' del DNA.
4. El estudio de la capacidad de la B35DNAP para replicar un DNA con sitios abásicos y la implicación del subdominio TPR2 en esta función.

Por otro lado, debido a la gran cantidad de ORFs en el genoma de Bam35 que codifican proteínas de función desconocida y a la necesidad de obtener más datos para entender su biología decidimos:

5. Realizar el estudio del interactoma proteico de Bam35 utilizando para ello el sistema de doble híbrido de levaduras.







## MATERIALES Y MÉTODOS

3





## 1. Nucleótidos

Los dNTPs no marcados radiactivamente fueron proporcionados por GE Healthcare. [ $\gamma$ - $^{32}\text{P}$ ]ATP (3.000 Ci/mmol), [ $\alpha$ - $^{32}\text{P}$ ]dATP (3.000 Ci/mmol), [ $\alpha$ - $^{32}\text{P}$ ]dTTP (3.000 Ci/mmol), [ $\alpha$ - $^{32}\text{P}$ ]dCTP (3.000 Ci/mmol) y [ $\alpha$ - $^{32}\text{P}$ ]dGTP (3.000 Ci/mmol) (1 Ci = 37 GBq) se compraron a PerkinElmer Inc.



## 2. Oligonucleótidos y DNA

Todos los oligonucleótidos utilizados fueron sintetizados por Sigma-Aldrich excepto los sustratos M33GFT, M33GFA y M33GTA (secuencias en **Tabla 4**) que fueron provistos por Isogen.

El genoma de Bam35 fue amablemente proporcionado por la Profesora Jaana Bamford (Universidad de Jyväskylä, Jyväskylä, Finlandia). El genoma circular de banda simple M13mp18 provino del stock del laboratorio. El *primer* universal M13 (5'-GTAAAACGACGGCCAGT-3') se obtuvo de Isogen. El plásmido pUC19 fue comprado a Thermo Scientific y, posteriormente se empleó para transformar *E. coli* XL-1 Blue y se aisló mediante el kit de extracción de DNA de Promega. Los *primers* de 6 nucleótidos (mer) con secuencia al azar protegidos de la actividad exonucleasa mediante enlaces fosforotioatos en el extremo 3' utilizados en el ensayo de amplificación por desplazamiento múltiple se obtuvieron de Integrated DNA Technologies (IDT).

El vector donador pDONR/Zeo para el sistema de doble híbrido en levaduras fue adquirido de Invitrogen, y los vectores de expresión pGADT7g, pGADCg, pGBKCg y pGBGT7g procedieron del stock del laboratorio del Dr. Uetz.



## 3. Proteínas

Las enzimas de restricción, la DNA ligasa T4, la polinucleótido quinasa T4, la DNAP Vent y el fragmento de Klenow se obtuvieron de New England Biolabs. La proteinasa K se compró a Affymetrix. La  $\Phi$ 29DNAP y su TP *wild type* provinieron del stock del laboratorio y fueron purificadas como se describe en Lázaro *et al.* (1995) y Mencía *et al.* (2011), respectivamente. La DNAP Kapa HiFi procedió de Kapa Biosystems.



## 4. DNA polimerasa *wild type* de Bam35 y variantes

### 4.1. Obtención de vectores de expresión

El genoma de Bam35 fue utilizado como molde para amplificar el gen 5 mediante la reacción en cadena de la polimerasa (PCR) utilizando la pareja de *primers* B35pol-fw y B35pol-rv (**Tabla 1**) y la DNAP Kapa HiFi en el buffer HiFi. El producto fue clonado en el vector pT7-4 entre los sitios de corte BamHI y HindIII (Tabor y Richardson, 1992) obteniendo el plásmido de expresión pT7-4::B35DNAP. Este vector se utilizó como

molde para obtener la variante de la B35DNAP D19A/E21A, que es deficiente para la actividad exonucleasa 3'-5' (B35DNAPexo-) y un mutante al que se le eliminó el subdominio TPR2 ( $\Delta$ TPR2) comprendido entre los aminoácidos E418 e Y458 mediante el kit de mutagénesis dirigida QuikChange (Agilent Technologies), utilizando los oligonucleótidos Exo-fw y Exo-rv y  $\Delta$ TPR2-fw and  $\Delta$ TPR2-rv (**Tabla 1**), respectivamente.

| Oligonucleótido  | Secuencia  |
|------------------|--|
| B35pol-fw        | 5'-CGCGAAGCTTAAAGGAGGAAGCATATGATGACTACTACTAATAGAAAAAAGCGT-3' |
| B35pol-rv        | 5'-CCGGATCCTTATTATAAGAACTTAATTCACCTAATAGTTCTTCAT-3'          |
| Exo-fw           | 5'-CTTTACGCTAGCTACTGCAACACGTGGTTTGGACGGCGCCGTGTTTCGAA-3'     |
| Exo-rv           | 5'-TTCGAAACACGGCGCCGTCCAAACCACGTGTTGCAGTAGCTAGCGTAAAG-3'     |
| $\Delta$ TPR2-fw | 5'-TGGAATTTTGCAATGCAGCGGAATATGCTAAGGCTGAATATATTCAAC-3'       |
| $\Delta$ TPR2-rv | 5'-GTTGAATATATTCAGCCTTAGCATATTCCCCTGCATTGCAAATTTTCCA-3'      |

**Tabla 1. Oligonucleótidos utilizados para la generación de los vectores de expresión de la B35DNAP y sus variantes.**

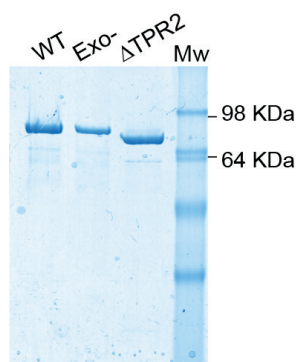
Se transformó *E. coli* XL-1 Blue, mediante choque térmico, con los vectores de expresión y, posteriormente, los genes de interés fueron completamente secuenciados.

#### 4.2. Expresión y purificación

Para expresar la B35DNAP *wild type* y sus variantes, B35DNAexo- y  $\Delta$ TPR2, se transformó *E. coli* BL21(DE3) con los plásmidos de expresión correspondientes (ver apartado 4.1, página 41). Los cultivos se crecieron en medio de autoinducción ZYM-5052 (Studier, 2005) con 100  $\mu$ g/mL de ampicilina durante 16-18 horas a 27 °C. Posteriormente, se recogieron las bacterias mediante centrifugación y se lisaron las células por abrasión con alúmina para luego resuspenderlas en tampón A (50 mM Tris-HCl pH 7,5, 1 mM EDTA, 7 mM  $\beta$ -mercaptoetanol y 5% (v/v) glicerol) ajustado a 0,5 M de NaCl. La alúmina y los restos bacterianos se eliminaron mediante centrifugación. A continuación, para precipitar el DNA, se ajustó la absorbancia a 260 nm a 120 U/mL, se añadió 0,3% (p/v) de polietilenimina y 1 M de NaCl y se centrifugó a 20.000 g durante 20 min. Al sobrenadante se le adicionó sulfato amónico hasta alcanzar una saturación del 69% y se volvió a centrifugar a 20.000 g durante 30 min.

Para purificar las proteínas, los precipitados se resuspendieron en tampón A con una fuerza iónica alrededor de 0,2 M de NaCl, se aplicaron a una columna Q-Sefarosa (GE Healthcare) y se eluyeron con tampón A a 0,25 M de NaCl. Los eluidos de la primera columna se pasaron por otra columna, en este caso de fosfocelulosa (P11, Whatman, Sigma) y después de extensivos lavados con tampón A a 0,25, 0,4 y 0,6 M de NaCl, las proteínas purificadas se eluyeron en el mismo tampón a 1 M de NaCl. Finalmente se adicionó Tween 20 y glicerol para unos porcentajes finales de 0,05 y 50% (v/v), respectivamente. La pureza de las proteínas (superior al 95%) se determinó mediante electroforesis en gel al 12% de poliacrilamida con dodecilsulfato sódico (12% SDS-PAGE) teñido con azul de Coomassie (**Figura 9**). Además, mediante ultracentrifugación en gradiente de glicerol se verificó el carácter intrínseco de la actividad de polimerización de las B35DNAPs

wild type y  $\Delta$ TPR2 (Figura S1 y Figura S2). Las proteínas se almacenaron a -20 °C para un uso inmediato o a -70 °C para largo plazo.



**Figura 9. Gel al 12% SDS-PAGE de la B35DNAP wild type, exo- y  $\Delta$ TPR2.** El gel muestra 1  $\mu$ g de cada proteína purificada.



## 5. Proteína terminal wild type de Bam35 y variantes

### 5.1. Obtención de vectores de expresión

El genoma de Bam35 fue utilizado como molde para amplificar el gen 4 mediante PCR utilizando la pareja de primers B35TP-fw y B35TP-rv (Tabla 2) y la DNAP Vent. El producto se clonó en el plásmido pT7-7 entre los sitios de corte NdeI y BamHI obteniendo el vector pT7-7::B35TP. Las variantes de la TP de Bam35 Y172F, Y172A, Y194F, Y194A, Y202F, Y215F, Y229F, Y234F, Y244F e Y245F se obtuvieron realizando mutagénesis dirigida sobre el plásmido pT7-7::B35TP utilizando los oligonucleótidos de la Tabla 2 mediante el kit de mutagénesis dirigida QuikChange (Agilent Technologies).

| Oligonucleótido | Secuencia   |
|-----------------|---|
| B35TP-fw        | 5'-CCGGCCCATATGGCAAATAAACGGTTAAAGAAGAACTAGAA-3'           |
| B35TP-rv        | 5'-GTCCCGGATCCTTATTAGTAGTAGTCATCATTATCCCACTTTC-3'         |
| YFPTP-fw        | 5'-GGAGAAATCATGGTTATGTAAAGATATTTGAAGGGCAAGGGAAAGAG-3'     |
| YFPTP-rv        | 5'-CTCTTTCCCTTGGCCTTCAAATATCTTTACATAACCATGATTTCTCC-3'     |
| Y172F-fw        | 5'-GGAGAAATCATGGTTATGTAAAGATATTTGAAGGGCAAGGGAAAGAG-3'     |
| Y172F-rv        | 5'-CTCTTTCCCTTGGCCTTCAAATATCTTTACATAACCATGATTTCTCC-3'     |
| Y172A-fw        | 5'-GGAGAAATCATGGTTATGTAAAGATAGCTGAAGGGCAAGGGAAAGAG-3'     |
| Y172A-rv        | 5'-CTCTTTCCCTTGGCCTTCAAGTATCTTTACATAACCATGATTTCTCC-3'     |
| Y194F-fw        | 5'-CGATTATGGTCGGTGTGTTTGATGTAAAAGAGCGTGACAAATATTTAAAGG-3' |
| Y194F-rv        | 5'-CCTTTAAATATTTGTCACGCTCTTTTACATCAAACACACCGACCATAATCG-3' |
| Y194A-fw        | 5'-CGATTATGGTCGGTGTGGCTGATGTAAAAGAGCGTGACAAATATTTAAAGG-3' |
| Y194A-rv        | 5'-CCTTTAAATATTTGTCACGCTCTTTTACATCAGCCACACCGACCATAATCG-3' |
| Y202F-fw        | 5'-GATGTAAAAGAGCGTGACAAATTTTAAAGGAATTAGTAGCTAA-3'         |
| Y202F-rv        | 5'-TTAGCTACTAATTCCTTTAAAAATTTGTCACGCTCTTTTACATC-3'        |
| Y215F-fw        | 5'-GTAGCTAAGCTACGTTCTTTACCTTTTGAAAAGGCTAAAAAGAAC-3'       |
| Y215F-rv        | 5'-GTTCTTTTGTAGCCTTTTCAAAAGGTAAAGAACGTAGCTTAGCTAC-3'      |
| Y229F-fw        | 5'-AACGCAAAGGAGATACAAAAGATATTTGATACAAAGTCGTATAAG-3'       |
| Y229F-rv        | 5'-CTTATACGACTTTGTATCAAATATCTTTTGTATCTCCTTTGCGTT-3'       |
| Y234F-fw        | 5'-AAGATATACGATACAAAAGTCGTTTAAAGAAAGAAAGTTGGGAT-3'        |
| Y234F-rv        | 5'-ATCCCAACTTTCTTTCTTAAACGACTTTGTATCGTATATCTT-3'          |
| Y244F-fw        | 5'-GAAAGTTGGGATAATGATGACTTTTACTAATAAGGATCCTCTAGAGT-3'     |
| Y244F-rv        | 5'-ACTCTAGAGGATCCTTATTAGTAAAAGTCATCATTATCCCACTTTC-3'      |
| Y245F-fw        | 5'-GAAAGTTGGGATAATGATGACTACTTTTAAATAAGGATCCTCTAGAGT-3'    |
| Y245F-rv        | 5'-ACTCTAGAGGATCCTTATTAAAGTAGTCATCATTATCCCACTTTC-3'       |

**Tabla 2. Oligonucleótidos utilizados para la generación de los vectores de expresión de la TP de Bam35 y sus variantes.**

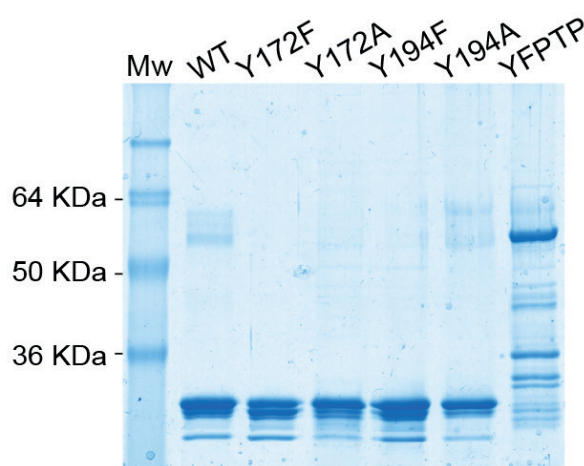
La proteína de fusión de la YFP a la parte N-terminal de la TP de Bam35 (YFPTP) se obtuvo amplificando el gen 4 con los *primers* YFPTP-fw e YFPTP-rv (**Tabla 2**) para, posteriormente, clonar el producto entre los sitios EcoRI y BamHI del plásmido pStrep-YFP generando el vector pStrep-YFP::B35TP (Redrejo-Rodríguez *et al.*, 2013a).

Se transformó *E. coli* XL-1 Blue con los vectores de expresión y posteriormente los genes de interés fueron completamente secuenciados.

### 5.2. Expresión y purificación

La expresión de la proteína codificada por el gen 4 del bacteriófago Bam35 (TP) y sus respectivos mutantes se realizó de manera similar a la B35DNAP (apartado 4.2, página 42). Una vez que se obtuvo el precipitado de sulfato amónico que contenía las proteínas de interés, se procedió a su purificación resuspendiéndolos en tampón A con una fuerza iónica alrededor de 0,4 M de NaCl, y se aplicaron a una columna Q-Sefarosa (GE Healthcare). Posteriormente, la fracción no unida se introdujo en una columna de fosfocelulosa (P11, Whatman, Sigma) y después de extensivos lavados con tampón A a 0,5, 0,6 y 0,7 M de NaCl las proteínas purificadas se eluyeron en el mismo tampón a 0,8 M de NaCl y se dializaron frente a otro tampón que contenía 50 mM Tris-HCl a pH 7,5, 1 mM EDTA, 7 mM  $\beta$ -mercaptoetanol, 50% (v/v) glicerol y 1 M NaCl.

Para expresar la proteína de fusión YFPTP de Bam35 se creció, en medio de Luria Bertani (LB) con 150  $\mu$ g/mL de ampicilina a 30 °C, un cultivo de *E. coli* BL21(DE3) transformado con el vector de expresión pStrep-YFP::B35TP (apartado 5.1, página 43). Cuando la densidad óptica a 600 nm ( $DO_{600}$ ) fue de 0,3 se indujo su expresión añadiendo 0,25 mM de isopropil  $\beta$ -D-tiogalactopiranosido (IPTG) y el cultivo se creció otros 30 min a 24 °C. Posteriormente, se purificó la proteína utilizando una resina Strep-Tactina (IBA GmbH) siguiendo las instrucciones del fabricante.



**Figura 10.** Gel al 12% SDS-PAGE de la TP *wild type* de Bam35 y sus variantes purificadas. El gel muestra 1  $\mu$ g de cada proteína purificada.

La pureza de las proteínas (superior al 95%) se analizó mediante electroforesis en gel al 12% SDS-PAGE teñido con azul de Coomassie (**Figura 10**). Las proteínas se almacenaron a -20 °C para un uso inmediato o a -70 °C para largo plazo.

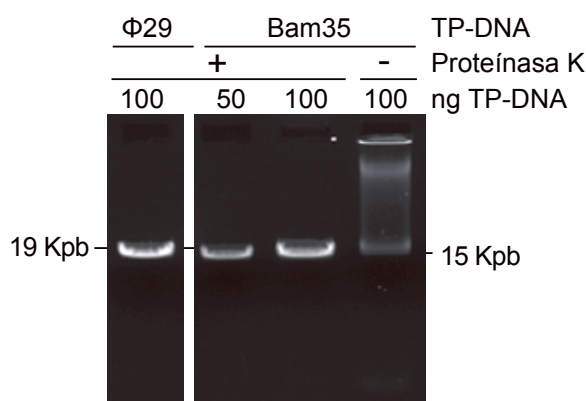


## 6. Obtención de TP-DNA de Bam35

Las partículas virales de Bam35 así como la cepa HER1410 de *B. thuringiensis* fueron amablemente proporcionados por la Profesora Jaana Bamford (Universidad de Jyväskylä, Jyväskylä, Finlandia).

Con el objetivo de obtener TP-DNA de Bam35 se crecieron, a 30 °C en LB, 500 mL de un cultivo de *B. thuringiensis* HER1410 hasta una densidad celular de  $3,6 \times 10^7$  unidades formadora de colonias por mL (ufc/mL). Posteriormente, se añadió  $MgCl_2$  a una concentración final de 10 mM y se infectó con el bacteriófago Bam35 a una multiplicidad de 1 unidad formadora de placa por bacteria (ufp/bacteria). Cuando el cultivo lisó, se centrifugó durante 10 min a 12.000 g para eliminar los restos celulares. Las partículas virales que se encontraban en el sobrenadante fueron precipitadas incubando con 6% (p/v) de polietilenglicol 6000 (Serva), durante 2 horas y media y después se centrifugaron durante 20 min a 25.000 g. El precipitado fue disuelto en 6 mL de tampón fosfato potásico a 10 mM (pH 7,2) y 10 mM de  $MgCl_2$  durante toda la noche a 4 °C. Posteriormente, la suspensión viral se centrifugó durante 20 min a 25.000 g y, para purificar parcialmente las partículas virales de Bam35, el sobrenadante se sometió, de manera consecutiva, a cromatografía en columna de fosfocelulosa (P11, Whatman) y Q-Sefarosa (GE Healthcare). Las partículas del bacteriófago Bam35 del eluido se concentraron volviendo a precipitar con polietilenglicol y centrifugación.

Para romper las partículas virales, el *pellet* se resuspendió en un tampón con 4 M de cloruro de guanidina, 25% (v/v) de glicerol, 200 mM de NaCl, 50 mM Tris-HCl pH 7,5, 10 mM de ditiotretitol (DTT), 1 mM de EDTA y 0,05% (v/v) Tween 20 (Sigma) y se incubó durante 90 min en hielo. Posteriormente se eliminó



**Figura 11. Resultado de la obtención del TP-DNA de Bam35.** El gel al 0,8% de agarosa muestra que el TP-DNA de Bam35 obtenido tiene el tamaño esperado (15 Kpb) además de una proteína covalentemente unida puesto que al no tratar con proteinasa K el genoma no se resuelve correctamente. El TP-DNA de Φ29 tratado con proteinasa K se utilizó de marcador de tamaño. Los ng de TP-DNA cargados en cada caso están indicados

el cloruro de guanidina mediante diálisis frente al mismo tampón sin dicho compuesto. El producto se clarificó por centrifugación y el TP-DNA de Bam35 obtenido se trató con 1 mg/mL de proteinasa K en presencia de 0,1% de SDS para, posteriormente analizarlo mediante electroforesis en gel de agarosa al 0,8% (McDonell *et al.*, 1977) seguido de tinción con bromuro de etidio (**Figura 11**). Además, se cargó una muestra sin tratar con proteinasa K para comprobar que la proteína covalentemente unida no se había perdido durante la purificación (carril de la derecha). Como marcador de tamaño se utilizó 100 ng de TP-DNA de  $\Phi$ 29 tratado con proteinasa K en las mismas condiciones que el TP-DNA de Bam35.



## 7. Ensayos con la TP y la B35DNAP

### 7.1. Ensayo de iniciación (formación del complejo TP-dNMP)

En la reacción estándar se preincubaron 12,5 nM (25 ng) de B35DNAP con 34 nM (25 ng) de TP o de YFPTP (50 ng) de Bam35 durante 10 min a temperatura ambiente, para favorecer la interacción entre las proteínas. Posteriormente, se añadió el complejo TP/B35DNAP a una mezcla de reacción de 25  $\mu$ L finales a 50 mM Tris-HCl (pH 7,5 en el caso de la activación con  $MgCl_2$  o 8,5 con  $MnCl_2$ , excepto que se indique lo contrario), 1 mM DTT, 4% (v/v) glicerol, 0,1 mg/mL de albúmina de suero bovino (BSA), 0,05% (v/v) de Tween 20, 0,2  $\mu$ M del dNTP indicado y 1  $\mu$ Ci del mismo [ $\alpha$ - $^{32}P$ ]dNTP para la iniciación independiente de molde ó 2  $\mu$ Ci cuando fue dependiente de molde.

Cuando el producto de la reacción de iniciación sin molde se trató con proteinasa K se adicionaron 0,5  $\mu$ g/ $\mu$ L de enzima y 0,1% de SDS a la mitad de la mezcla de reacción y se incubó durante 20 min a 37 °C. Para analizar la dependencia de pH en la reacción de iniciación, los pHs 7, 7,5, 8 y 8,5 se obtuvieron con una solución tampón Tris-HCl mientras que para 9 y 9,5 se utilizó CHES-NaOH.

En el caso de la iniciación dependiente de molde, se adicionó 315 nM de un oligonucleótido de cadena sencilla protegido, mediante dos enlaces fosforotioatos en el extremo 3', de la actividad exonucleasa de la

| Oligonucleótido  | Secuencia                             |
|------------------|---------------------------------------|
| WT               | 5'-TTTAATAGGTTGGTAGGGGTACCATAA*T*A-3' |
| 1C (CTAATA)      | 5'-TTTAATAGGTTGGTAGGGGTACCATAA*T*C-3' |
| 2C (ACAATA)      | 5'-TTTAATAGGTTGGTAGGGGTACCATAA*C*A-3' |
| 3C (ATCATA)      | 5'-TTTAATAGGTTGGTAGGGGTACCATAC*T*A-3' |
| 4C (ATACTA)      | 5'-TTTAATAGGTTGGTAGGGGTACCATCA*T*A-3' |
| 5C (ATAACA)      | 5'-TTTAATAGGTTGGTAGGGGTACCACAA*T*A-3' |
| 1C-3C (CTCATA)   | 5'-TTTAATAGGTTGGTAGGGGTACCATAC*T*C-3' |
| 29R- $\Phi$ 29fw | 5'-TGGTGTATGTTGTCGCTGTACCCTACT*T*T-3' |
| 29R- $\Phi$ 29rv | 5'-AAAGTAGGGTACAGCGACAACATACAC*C*A-3' |

**Tabla 3. Oligonucleótidos utilizados en el ensayo de iniciación.** Los \* indican la posición de los enlaces fosforotioatos. El nombre de los oligonucleótidos indican la posición desde el extremo 3' en la que se ha modificado la secuencia *wild type* del origen izquierdo del genoma de Bam35 por una citosina.



DNAP (ver secuencias en **Tabla 3**) o 75 ng de TP-DNA de Bam35 (ver apartado 6, página 45).

Todas las reacciones fueron activadas con 1 mM  $\text{MnCl}_2$  o 10 mM  $\text{MgCl}_2$ . Excepto cuando se indica, las muestras fueron incubadas durante 30 min a 37 °C.

Cuando se utilizó  $\Phi 29$  como control, la reacción se llevó a cabo como describe Gella *et al.* (2014) con 4  $\mu\text{Ci}$  de  $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ , 1 mM de  $\text{MnCl}_2$  y 315 nM de un oligonucleótido de 29-mer de doble cadena con la secuencia del origen derecho del genoma de  $\Phi 29$  (**Tabla 3**). Las incubaciones fueron de 30 min a 30 °C.

En todos los casos las reacciones se pararon con 10 mM EDTA y 0,1% SDS (concentraciones finales). Los productos se analizaron en gel MiniProtean III (BioRad) al 12% SDS-PAGE usando un marcador de peso molecular preteñido (SeeBlue® Plus2, Invitrogen) seguido de autorradiografía y, cuando se indica, de cuantificación con el programa *ImageJ* (Rasband, 1997-2016).

### 7.2. Mapeo del residuo iniciador de la TP de Bam35

Para determinar si el aminoácido de la TP al que la B35DNAP une el primer nucleótido de la replicación era una serina, treonina o tirosina, se realizó un ensayo de iniciación sin molde como se describe en el apartado 7.1 (página 46) pero con 4  $\mu\text{Ci}$  de  $[\alpha\text{-}^{32}\text{P}]\text{dATP}$  y 1 mM de  $\text{MnCl}_2$  y se llevó como control  $\Phi 29$  sin modificar las condiciones. Posteriormente, los productos fueron sometidos a tratamiento alcalino con 100 mM NaOH durante 6 min a 95 °C después de los cuales se neutralizaron con 100 mM HCl y se analizaron en gel MiniProtean III (BioRad) al 12% SDS-PAGE.

La posición del residuo iniciador fue determinada mediante la fragmentación de la proteína con bromuro de cianógeno (CNBr, sigma). Para ello, se realizó un ensayo de iniciación sin molde con 1 mM de  $\text{MnCl}_2$  y 272 nM de TP de Bam35. La incubación se realizó a 37 °C toda la noche (18 h) para incrementar la cantidad de producto. A continuación, las muestras se trataron con 1,2 mM de CNBr en presencia de 200 mM HCl durante 20 horas a temperatura ambiente (Andreev *et al.*, 2010). Finalmente, se neutralizó la reacción con 200 mM NaOH y los productos se analizaron en gel MiniProtean III (BioRad) al 18% SDS-PAGE.

### 7.3. Ensayo de iniciación con extractos bacterianos

Para realizar este ensayo, se crecieron los cultivos de *E. coli* BL21(DE3) transformados con el plásmido vacío pT7-7 o con los plásmidos pT7-7::B35TP para la expresión de las proteínas *wild type* o sus variantes Y194F, Y202F, Y215F, Y229F, Y234F, Y244F o Y245F durante 18 horas a 27 °C en medio de autoinducción ZYM-5052 con 100  $\mu\text{g}/\text{mL}$  de ampicilina. Posteriormente, se recogieron las bacterias de 1 mL de cada cultivo mediante centrifugación y se resuspendieron en 0,5 mL de tampón A (composición en apartado 4.2, página 42) con 0,5 M NaCl. Se lisaron las células por sonicación y se clarificó el extracto por centrifugación a 4 °C durante 10 min a 20.000 g. Los sobrenadantes que contenían las proteínas solubles, incluidas las TPs,

fueron diluidas 1:2 con glicerol 100%. Finalmente, se llevó a cabo un ensayo de iniciación sin molde con 1 mM de  $\text{MnCl}_2$  y con 0,5 o 2  $\mu\text{L}$  de las diluciones de los extractos en glicerol.

#### **7.4. Análisis de la interacción TP/B35DNAP a través de un ensayo funcional**

Para comparar la interacción funcional entre la TP *wild type* y sus variantes con la B35DNAP, analizamos la capacidad de estas TPs para inhibir la deoxiadenucleación de la proteína de fusión YFPTP. Para ello, se realizó un ensayo de iniciación sin molde con 1 mM de  $\text{MnCl}_2$  y diferentes concentraciones de las TPs. Después de 2,5 min de incubación, se adicionó YFPTP para una concentración final de 34 nM y se incubó durante otros 2,5 min. Además, se ajustó la cantidad de proteína total en todas las muestras adicionando BSA. Las reacciones fueron paradas y procesadas como se describe en el apartado 7.1 (página 46). Los productos marcados radiativamente fueron detectados mediante autorradiografía y cuantificados con el *ImageJ* (Rasband, 1997-2016).

#### **7.5. Replicación truncada y completa de oligonucleótidos de banda simple**

Las reacciones se realizaron como se describe para la iniciación con molde con  $\text{MnCl}_2$  (apartado 7.1, página 46) pero con 5  $\mu\text{M}$  de los dNTPs indicados, 2  $\mu\text{Ci}$  de  $[\alpha\text{-}^{32}\text{P}]\text{dNTP}$  complementario a la tercera base del molde utilizado (dNTP iniciador) y, cuando se indica, 100  $\mu\text{M}$  del didesoxinucleótido (ddNTP) correspondiente para terminar la replicación en las posiciones indicadas. Para poder analizar todos los productos de replicación generados, las muestras se corrieron en un gel de alta resolución al 15% SDS-PAGE (245 x 300 x 0,5 mm) y el resultado se detectó mediante autorradiografía.

#### **7.6. Replicación de TP-DNA**

Se incubaron 12,5 nM de B35DNAP y 133 nM de TP de Bam35 durante 10 min a temperatura ambiente antes de ser añadidas a una mezcla que contenía, en 25  $\mu\text{L}$ , 50 mM Tris-HCl a pH 7,5, 1 mM DTT, 4% glicerol, 0,1 mg/mL de BSA, 0,05% Tween 20, 150 mM NaCl, 20 mM sulfato amónico, 20  $\mu\text{M}$  de cada uno de los cuatro dNTPs, 1  $\mu\text{Ci}$  de  $[\alpha\text{-}^{32}\text{P}]\text{dATP}$  y 100 ng de TP-DNA de Bam35 (obtenido como indica el apartado 6, página 45). Las reacciones fueron activadas con 1 mM  $\text{MnCl}_2$  y/o 10 mM  $\text{MgCl}_2$  e incubadas durante los tiempos indicados. Posteriormente, se pararon las reacciones con 10 mM EDTA y 0,1% SDS (concentraciones finales) y se pasaron por columnas de Sephadex G-50 con 0,1% SDS para eliminar el nucleótido radiactivo no incorporado en la replicación. Los resultados se analizaron mediante electroforesis en gel de agarosa alcalino al 0,7% y se detectaron por autorradiografía. Como indicador de tamaño se utilizó el DNA de  $\lambda$  digerido con HindIII y marcado con el fragmento de Klenow en presencia de  $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ , como indica Sambrook y Russell (2001).



## 8. Ensayos con la B35DNAP

### 8.1. Replicación y degradación de oligonucleótidos

Los oligonucleótidos que se utilizaron como *primer* (**Tabla 4**) fueron purificados a partir de geles al 20% de poliacrilamida y 8 M urea. Una vez purificados fueron marcados en el extremo 5' con [ $\gamma$ - $^{32}$ P]ATP usando la polinucleótido kinasa de T4 e hibridados con un exceso de 1,2 molar de un oligonucleótido molde complementario no marcado (**Tabla 4**) en presencia de 50 mM NaCl y 50 mM Tris-HCl, pH 7,5. Los sustratos *primer*/molde obtenidos se utilizaron para estudiar las actividades polimerasa y exonucleasa de la B35DNAP además de su capacidad para replicar a través de daños en el DNA molde.

|            | Oligonucleótido | Secuencia  |
|------------|-----------------|--|
| PRIMER (C) | C15 (15-mer)    | 5'-GATCACAGTGAGTAC-3'                                  |
|            | C15G (15-mer)   | 5'-GATCACAGTGAGTAG-3'                                  |
|            | C16 (16-mer)    | 5'-GATCACAGTGAGTACA-3'                                 |
|            | C19 (19-mer)    | 5'-GATCACAGTGAGTACAATA-3'                              |
|            | C21 (21-mer)    | 5'-GATCACAGTGAGTACAATAGA-3'                            |
|            | C22 (22-mer)    | 5'-GATCACAGTGAGTACAATAGAA-3'                           |
|            | C23 (23-mer)    | 5'-GATCACAGTGAGTACAATAGAAC-3'                          |
|            | C25 (25-mer)    | 5'-GATCACAGTGAGTACAATAGAACGA-3'                        |
| MOLDE (M)  | M21 (21-mer)    | 5'-TCTATTGTACTCACTGTGATC-3'                            |
|            | M23 (23-mer)    | 5'-GTTCTATTGTACTCACTGTGATC-3'                          |
|            | M25 (25-mer)    | 5'-TCGTTCTATTGTACTCACTGTGATC-3'                        |
|            | M27 (27-mer)    | 5'-CGTCGTTCTATTGTACTCACTGTGATC-3'                      |
|            | M29 (29-mer)    | 5'-GCCGTCGTTCTATTGTACTCACTGTGATC-3'                    |
|            | M33GTT (33-mer) | 5'-ACTGGCCGTCGTTCTATTGTACTCACTGTGATC-3'                |
|            | M33GFT (33-mer) | 5'-ACTGGCCGTCGTTCTATF <sup>15</sup> GTACTCACTGTGATC-3' |
|            | M33GTA (33-mer) | 5'-ACTGGCCGTCGTTCTAAT <sup>16</sup> GTACTCACTGTGATC-3' |
|            | M33GFA (33-mer) | 5'-ACTGGCCGTCGTTCTAA <sup>16</sup> FGTACTCACTGTGATC-3' |
|            | M35 (35-mer)    | 5'-CAACTGGCCGTCGTTCTATTGTACTCACTGTGATC-3'              |

**Tabla 4. Oligonucleótidos utilizados en ensayos de replicación utilizando DNA como *primer*.** Los oligonucleótidos *primers* (C) fueron hibridados con distintos moldes (M) obteniendo los sustratos *primer*/molde indicados en resultados. Los números en los nombres indican el tamaño del oligonucleótido en mer. En el caso del C15G indica que el nucleótido del extremo 3' es una G en lugar de C como en el oligonucleótido C15. En los moldes M33GTT, M33GFT, M33GTA y M33GFA las letras indican los nucleótidos en sentido 3'-5' de las posiciones 15, 16 y 17. La F indica la presencia de un tetrahidrofurano (THF), un análogo de sitios abásicos.

Los ensayos se llevaron a cabo en un volumen final de 20  $\mu$ L que contenía 50 mM Tris-HCl a pH 7,5, 1 mM DTT, 4% (v/v) de glicerol, 0,1 mg/mL de BSA y, al menos que se indique lo contrario, 1 nM de la hibridación del oligonucleótido molde con el *primer* marcado en 5', 10 nM de la DNAP correspondiente o 200 nM en el caso de  $\Delta$ TPR2 y la concentración de dNTPs indicada. Las reacciones se activaron con 10 mM de  $MgCl_2$  y después de la incubación a 37 °C durante los tiempos indicados, se pararon con 10  $\mu$ L de tampón de carga con formamida (98% formamida, 20 mM EDTA, 0,5% de azul de bromofenol y 0,5% de xileno cianol). Finalmente, las muestras se analizaron en geles al 20% poliacrilamida, 8 M urea y los productos se detectaron por autorradiografía. Alternativamente, para la cuantificación de los experimentos se utilizaron imágenes digitalizadas mediante el equipo *Fujifilm BAS1000 phosphorimager*

que se densitometraron con el programa *ImageJ* (Rasband, 1997-2016). Estos resultados se utilizaron para analizar con más detalle la replicación de oligonucleótidos con THF por la B35DNAP siguiendo las siguientes ecuaciones (Kokoska *et al.*, 2003):

Las bandas 16 y 33 mer corresponden con los productos de polimerización obtenidos

$$\text{Probabilidad de inserción de un nucleótido} \\ \text{enfrente del THF} = \frac{\text{Intensidad de las bandas (16 mer+33 mer)}}{\text{Suma de las intensidades de todas las bandas}} \times 100$$

$$\text{Probabilidad de la replicación completa} \\ \text{del molde con o sin THF} = \frac{\text{Intensidad de la banda 33 mer}}{\text{Suma de las intensidades de todas las bandas}} \times 100$$

$$\text{Eficiencia relativa de la} \\ \text{replicación del sustrato con} \\ \text{THF con respecto a sin THF} = \frac{\text{Probabilidad de la replicación completa del molde con THF}}{\text{Probabilidad de la replicación completa del molde sin THF}} \times 100$$

### 8.2. Retraso en gel

Para analizar la interacción entre la B35DNAP con los diferentes sustratos *primer*/molde, se llevó a cabo un ensayo de cambio de movilidad electroforética o retraso en gel de dichos sustratos. Para ello, se realizó una incubación en hielo durante 10 min de 5 nM de B35DNAP con 2 nM de los distintos sustratos *primer*/molde (marcado en el 5' del *primer*) en un tampón con 50 mM Tris-HCl, pH 7,5, 20 mM de sulfato amónico, 1 mM DTT, 4% (v/v) de glicerol, 0,1 mg/mL de BSA, 0,05% (v/v) Tween 20 y 10 mM de MgCl<sub>2</sub>. Posteriormente, las muestras fueron sometidas a electroforesis en gel al 4% de poliacrilamida (80:1, monómero/bis) con 12 mM de Tris-Acetato (pH 7,5) y 1 mM de EDTA. La electroforesis se desarrolló a temperatura ambiente en el mismo tampón a 20 mA como describe Carthew *et al.* (1985). El complejo B35DNAP-DNA se detecta como un retraso en la migración del DNA marcado.

### 8.3. Replicación del DNA circular de M13

Este ensayo se utilizó para estudiar la procesividad y la capacidad de desplazamiento de banda de la B35DNAP.

El DNA circular de cadena simple M13mp18 fue hibridado con el *primer* universal M13 (PU) (5'-GTAAAACGACGGCCAGT-3') en presencia de 0,2 M de NaCl y 60 mM Tris-HCl, pH 7,5 obteniendo el molde PU/M13. Las reacciones se llevaron a cabo en 25 µL finales con 50 mM Tris-HCl, pH 7,5, 1 mM DTT, 4% (v/v) de glicerol, 0,1 mg/mL de BSA, 0,05% (v/v) Tween 20, 40 µM de dNTPs, 0,5 µCi [α-<sup>32</sup>P]dATP, 3,2 nM del molde PU/M13, 10 mM MgCl<sub>2</sub> y la concentración indicada de DNAP. Después de una incubación durante 20 min a 37 °C las reacciones se pararon añadiendo 30 mM de EDTA, 0,5% de SDS y 0,3 M de NaCl.

Los productos de la replicación del DNA se analizaron en gel alcalino al 0,7% de agarosa. Los geles fueron teñidos con bromuro de etidio, secados y sometidos a autorradiografía.

Como indicador de tamaño se marcó el DNA de  $\lambda$  (digerido con HindIII) con el fragmento de Klenow en presencia de  $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ , como indica Sambrook y Russell (2001).

#### 8.4. Amplificación por desplazamiento múltiple (MDA) de un plásmido de DNA

La MDA consiste en la amplificación isotérmica de DNA utilizando *primers* de 6 a 8-mer con secuencia al azar. Este ensayo se realiza con una DNAP con capacidad de desplazamiento de banda y alta procesividad de manera que al elongar los *primers* va generando cadenas de DNA de banda simple laterales que son susceptibles de hibridar con más *primers* y por lo tanto, de amplificarse (Dean *et al.*, 2002) (**Figura 19**, página 68).

En esta tesis decidimos amplificar el vector pUC19 porque posee el gen lacZ $\alpha$  (inducible por IPTG) que fue utilizado como gen indicador para realizar el posterior análisis de la fidelidad de la B35DNAP (ver 8.5, página 51).

El ensayo se realizó de manera similar a la descrita previamente por Nelson *et al.* (2002). Las reacciones se llevaron a cabo en 20  $\mu\text{L}$  de una mezcla 40 mM Tris-HCl, pH 7,5, 50 mM KCl, 0,025% (v/v) Tween 20, 10 mM  $\text{MgCl}_2$ , 500  $\mu\text{M}$  de dNTPs, 25  $\mu\text{M}$  de *primer* de 6 o 8-mer con secuencia al azar protegidos de la actividad exonucleasa mediante enlaces fosforotioatos en el extremo 3', 50 nM de B35DNAP y la cantidad indicada de pUC19. La incubación se realizó durante 16 horas a 37 °C o a las temperaturas indicadas. Finalmente, para obtener el tamaño unidad del plásmido y poder analizar el resultado, se tomaron 2  $\mu\text{L}$  de cada reacción y se digirieron con la enzima EcoRI-HF durante 2 horas a 37 °C antes de someterlas a electroforesis en gel al 0,7% de agarosa.

#### 8.5. Análisis de la fidelidad de la B35DNAP

Para analizar la fidelidad de la B35DNAP, se tomó una fracción del producto de la reacción del ensayo de MDA (ver 8.4, página 51) y se digirió, durante 2 horas a 37 °C, con EcoRI-HF y DpnI para obtener el tamaño unidad del plásmido y además eliminar el pUC19 molde, purificado de la bacteria y, por tanto, metilado. El producto digerido se purificó con el kit *clean-up* de PCR de Qiagen y, se estimó la concentración a partir de la medida de absorbancia a 260 nm en un espectrofotómetro Nanodrop. A continuación, se re-circularizaron 190 ng del pUC19 digerido con 400 unidades de DNA ligasa de T4 en un volumen final de 100  $\mu\text{L}$ . La ligación se confirmó mediante electroforesis en gel de agarosa al 0,7%. Posteriormente, se transformó *E. coli* XL-1 Blue con 20 ng del pUC19 religado y se plaqueó en LB agar con 150  $\mu\text{g}/\text{mL}$  de ampicilina, 40  $\mu\text{g}/\text{mL}$  de

5-bromo-4-cloro-3-indolil- $\beta$ -D-galactopiranosido (x-gal) y 0,5 mM de IPTG para inducir la expresión del gen lacZ $\alpha$  y poder distinguir las colonias con mutaciones en dicho gen (colonias blancas) de las que no tienen mutaciones o tienen mutaciones silenciosas (colonias azules). Después de incubar las placas toda la noche a 37 °C se contaron ambos tipos de colonias y se estimó la tasa de error cometida por la B35DNAP. Además, debido a la posibilidad de mutaciones generadas por las bacterias durante la replicación o reparación de plásmidos (Jozwiakowski y Connolly, 2009), la frecuencia de estas mutaciones, correspondientes al fondo, fue determinada en cada experimento realizando el mismo protocolo (excepto la adición de DpnI durante la digestión) con el plásmido pUC19 procedente de la bacteria.

La fórmula utilizada para hallar la tasa de error (E) de la B35DNAP fue:

$$E = \frac{F - F_0}{pb \times d}$$

Donde:

$$\text{La frecuencia de mutación: } F = \frac{\text{Número de colonias blancas}}{\text{Número de colonias totales}}$$

$$\text{El fondo de la frecuencia de mutación: } F_0 = \frac{\text{Número de colonias blancas con pUC19 original}}{\text{Número de colonias totales con pUC19 original}}$$

Número de pares de bases detectables: pb= 342 pb (tamaño del gen lacZ $\alpha$ )

$$\text{Número de duplicaciones del molde pUC19: } d = \log_2 \frac{\text{ng de pUC19 finales (amplificado)}}{\text{ng de pUC19 iniciales (añadido)}}$$

Para analizar los tipos de mutación más comunes producidos por la B35DNAP, se secuenciaron 43 colonias blancas con el *primer* LacZfw (5'-AGCTTGTCTGTAAGCGGATGCCG-3').



## 9. Sistema de doble híbrido de levaduras

Con el objetivo de poder predecir la función de alguna de las proteínas de función desconocida codificadas por el genoma de Bam35, decidimos analizar las interacciones entre esas proteínas mediante el sistema de doble híbrido de levaduras.

### 9.1. Clonaje de los ORFs de Bam35 en el vector de entrada pDONR/Zeo

La secuencia del DNA de Bam35 se obtuvo de la base de datos de genomas NCBI (NC\_005258.1) y los *primers* se diseñaron, utilizando el programa *Geneious* (Kearse *et al.*, 2012), sin codón de parada y con una longitud de 20-30 nucleótidos complementarios al ORF que se quería amplificar (**Tabla 5**). A los *primers forward* se les añadió un fragmento de la secuencia attB1 (5'-AAAAAGCAGGCTTA-3') en el extremo 5' mientras que a



los *reverse* se les adicionó la secuencia attB2 (5'-AGAAAGCTGGGTG-3') en el mismo extremo. Además, se predijeron los dominios de transmembrana de las proteínas codificadas por los ORFs mediante el servidor TMHMM (Krogh *et al.*, 2001) y se diseñaron oligonucleótidos para amplificar variantes de esos ORFs sin dichos dominios.

| ORF                   | Primer  | Secuencia  |
|-----------------------|---------|--|
| 1                     | Forward | 5'-AAAAAGCAGGCTTAATGAGTAACTACTGACTGCGCAAGAGGTTG-3'     |
|                       | Reverse | 5'-AGAAAGCTGGGTGATTGTCCAGTTTATTTTATTTTCCAATAGTCCTTT-3' |
| 2                     | Forward | 5'-AAAAAGCAGGCTTAATGGAACAATACTAAACAATTAAATAATGAGGTT-3' |
|                       | Reverse | 5'-AGAAAGCTGGGTGGCCTTGTAGTTGAAGTGTAAATGTTTTGAA-3'      |
| 3                     | Forward | 5'-AAAAAGCAGGCTTAATGAACGCTATTAATAATTTGGTTCAGGTATC-3'   |
|                       | Reverse | 5'-AGAAAGCTGGGTGCATGGTTTTCTTTCTGTAAATAGAGTTCATAAAG-3'  |
| 4                     | Forward | 5'-AAAAAGCAGGCTTAATGGCAAATAAACGGTTAAAGAAGAACTAG-3'     |
|                       | Reverse | 5'-AGAAAGCTGGGTGGTAGTAGTCATCATATCCCAACTTTCTTTCT-3'     |
| 5                     | Forward | 5'-AAAAAGCAGGCTTAATGATGACTACTACTAATAGAAAAAGCGTAGAG-3'  |
|                       | Reverse | 5'-AGAAAGCTGGGTGTAAGAACTTAATTCACCTAATAGTTCTTTTCATGT-3' |
| 6                     | Forward | 5'-AAAAAGCAGGCTTAATGTTGACGCCAAGGGAACAGGATACGTTA-3'     |
|                       | Reverse | 5'-AGAAAGCTGGGTGTAACAACCTGGATGGCTCTAGGTCTTTGATG-3'     |
| 7                     | Forward | 5'-AAAAAGCAGGCTTAATGCGTGACAAATTGCTCGACTTTATCATC-3'     |
|                       | Reverse | 5'-AGAAAGCTGGGTGGTCATCCTTCTTCCCTCCACATTCTTTTC-3'       |
| 8                     | Forward | 5'-AAAAAGCAGGCTTAATGTGGAGGGGAAGAAGGATGACTGATAAA-3'     |
|                       | Reverse | 5'-AGAAAGCTGGGTGGTCCTTTTCTACTATTCTTGTTCACAA-3'         |
| 9                     | Forward | 5'-AAAAAGCAGGCTTAATGGATGCTTTAACGGAACATCATAGCT-3'       |
|                       | Reverse | 5'-AGAAAGCTGGGTGATTGATTTGTATAAATCAGAAATCTAGTTTCTAT-3'  |
| 10                    | Forward | 5'-AAAAAGCAGGCTTAATGGAAGAAGAAAAGCATGTGAAGACGAA-3'      |
|                       | Reverse | 5'-AGAAAGCTGGGTGCAGGAACCAACTTAGGAAGCTTTTCT-3'          |
| 11                    | Forward | 5'-AAAAAGCAGGCTTAATGGGCAAAGACATTCATGTCGACGATACAG-3'    |
|                       | Reverse | 5'-AGAAAGCTGGGTGAAATGGAGCGCTTGCCGGAATGAGT-3'           |
| 12                    | Forward | 5'-AAAAAGCAGGCTTAATGCTAAACCAATTCAACCTACCGGAAAC-3'      |
|                       | Reverse | 5'-AGAAAGCTGGGTGCTCGCTCGGATAAGTACCGTATTGGACATAG-3'     |
| 13                    | Forward | 5'-AAAAAGCAGGCTTAATGTCCAATACGGTACTTATCCGAGCGAGT-3'     |
|                       | Reverse | 5'-AGAAAGCTGGGTGGTCGGGATGCCATTCATGTTTCTTCACACC-3'      |
| 14                    | Forward | 5'-AAAAAGCAGGCTTAATGAATGGCATCCGACTGACGAACATGTC-3'      |
|                       | Reverse | 5'-AGAAAGCTGGGTGAAGTTTCAAAGTGGCGCGAACAGGTTT-3'         |
| 15                    | Forward | 5'-AAAAAGCAGGCTTAATGGAGGGTAAATTTGCAGGGGTGCGACTCA-3'    |
|                       | Reverse | 5'-AGAAAGCTGGGTGTACGGCTTGACAACCTTCACTAAGTCCTTT-3'      |
| 16                    | Forward | 5'-AAAAAGCAGGCTTAATGGGTAACTATTTAGCCCGTCTTGGTGGG-3'     |
|                       | Reverse | 5'-AGAAAGCTGGGTGTACCGCTCCGCAATCGGTCTTACAAC-3'          |
| 17                    | Forward | 5'-AAAAAGCAGGCTTAATGCACGTAGTTGATATTTAGATGTAAATATGAA-3' |
|                       | Reverse | 5'-AGAAAGCTGGGTGGTTTGTTCCTCCTTCAGCTTGTGGGC-3'          |
| 18                    | Forward | 5'-AAAAAGCAGGCTTAATGGCACAACAGAAAGAGCAATATTCACCGCA-3'   |
|                       | Reverse | 5'-AGAAAGCTGGGTGCATTTGTAATCGTGATAATGTTTCTGATACTACA-3'  |
| 19                    | Forward | 5'-AAAAAGCAGGCTTAATGGGACTGAATCATTACTATCATGATTTATACC-3' |
|                       | Reverse | 5'-AGAAAGCTGGGTGCAACCTCCCCGCCACATAACCGAATGCAAT-3'      |
| 19'<br>(sin aa 52-76) | Forward | 5'-AAAAAGCAGGCTTAATGGGACTGAATCATTACTATCATGATTTATACC-3' |
|                       | Reverse | 5'-AGAAAGCTGGGTGTGTTACTGGGTCACTCTTGGCTTTACGTCTGTCTT-3' |
| 20                    | Forward | 5'-AAAAAGCAGGCTTAATGGACGCTCAAATGGTAGGTCAATTAATT-3'     |
|                       | Reverse | 5'-AGAAAGCTGGGTGCTCATGCGCCTTACCACCCATTTGCTCTAT-3'      |
| 20'<br>(sin aa 1-26)  | Forward | 5'-AAAAAGCAGGCTTAATGAAAAAGACACTAGAAGGCGTTACACAATCT-3'  |
|                       | Reverse | 5'-AGAAAGCTGGGTGCTCATGCGCCTTACCACCCATTTGCTCTAT-3'      |

| ORF                     | Primer  | Secuencia   |
|-------------------------|---------|---|
| 21                      | Forward | 5'-AAAAAGCAGGCTTAATGAACAACGACGCAAAGGGTTTAACG-3'           |
|                         | Reverse | 5'-AGAAAGCTGGGTGTTTAAACCACCCTTTCACCGTACCCAC-3'            |
| 21'<br>(sin aa 1-26)    | Forward | 5'-AAAAAGCAGGCTTAATG GAAAGAAACGTCTATCAAAATTGGCGCAA-3'     |
|                         | Reverse | 5'-AGAAAGCTGGGTGTTTAAACCACCCTTTCACCGTACCCAC-3'            |
| 22                      | Forward | 5'-AAAAAGCAGGCTTAATGAAAGAATTTACGGAGTCAATGCCATTTCGTC-3'    |
|                         | Reverse | 5'-AGAAAGCTGGGTGTACTCTCACCTCATTTAATAGTTTTATAAACTTGTGTC-3' |
| 23                      | Forward | 5'-AAAAAGCAGGCTTAATGACAAAAGTATTCGATATGATCGCGGGA-3'        |
|                         | Reverse | 5'-AGAAAGCTGGGTGTCGTCCTTGTATGTTTAAATGCCCGCAAT-3'          |
| 24                      | Forward | 5'-AAAAAGCAGGCTTAATGGATTACAAGCGCGACAATTTAAACCGT-3'        |
|                         | Reverse | 5'-AGAAAGCTGGGTGCACCCCATTAATGGATTTCCGAAATAATA-3'          |
| 25                      | Forward | 5'-AAAAAGCAGGCTTAATGCCAGAATTTGAAAGACTCGACCCTCT-3'         |
|                         | Reverse | 5'-AGAAAGCTGGGTGATAACTCCTAGAATGTGCGCCATAACTTG-3'          |
| 25'<br>(sin aa 1-44)    | Forward | 5'-AAAAAGCAGGCTTAATGAACAAGGCGAAACATAGCGGTGGTT-3'          |
|                         | Reverse | 5'-AGAAAGCTGGGTGATAACTCCTAGAATGTGCGCCATAACTTG-3'          |
| 26                      | Forward | 5'-AAAAAGCAGGCTTAATGGCTTTAGATGTTAGACCATTTCATAAAC-3'       |
|                         | Reverse | 5'-AGAAAGCTGGGTGTATACTTGCTTCAACTTGTCGTTTAGTTTG-3'         |
| 26'<br>(sin aa 177-250) | Forward | 5'-AAAAAGCAGGCTTAATGGCTTTAGATGTTAGACCATTTCATAAAC-3'       |
|                         | Reverse | 5'-AGAAAGCTGGGTGACCATTTAGCGACGCTTGTGAATATTCCACG-3'        |
| 27                      | Forward | 5'-AAAAAGCAGGCTTAATGGCGGAACCTACAGGGACTCAAACATAAAC-3'      |
|                         | Reverse | 5'-AGAAAGCTGGGTGTAGTAGTTTTACTTTTCATATCTGGGAAGGC-3'        |
| 27'<br>(sin aa 140-170) | Forward | 5'-AAAAAGCAGGCTTAATGGCGGAACCTACAGGGACTCAAACATAAAC-3'      |
|                         | Reverse | 5'-AGAAAGCTGGGTGTCCCATAAATGCACTTGTAAATAATCCTTTGTC-3'      |
| 28                      | Forward | 5'-AAAAAGCAGGCTTAATGACGAAGGATAATACATGGCGTGAGGTC-3'        |
|                         | Reverse | 5'-AGAAAGCTGGGTGCACCCCAAAATACGGACAGTTCC-3'                |
| 29                      | Forward | 5'-AAAAAGCAGGCTTAATGGCTTTGATTAAACCGTTTCCAGTAATA-3'        |
|                         | Reverse | 5'-AGAAAGCTGGGTGATATGCACATTGCCATGTGCCACCGATTTT-3'         |
| 30                      | Forward | 5'-AAAAAGCAGGCTTAATGGGTGTAATAGCTGATTATCACATCAT-3'         |
|                         | Reverse | 5'-AGAAAGCTGGGTGAATATTCACCTTCCACATTCCGTAACCCG-3'          |
| 31                      | Forward | 5'-AAAAAGCAGGCTTAATGGTGTACCATTTCGTAACCTATGCGTTTTT-3'      |
|                         | Reverse | 5'-AGAAAGCTGGGTGTATCAGCAAACCATATTGTGCGATCCTTGT-3'         |
| 32                      | Forward | 5'-AAAAAGCAGGCTTAATGGACGGTATATCGGGTGACGTAGATTTAA-3'       |
|                         | Reverse | 5'-AGAAAGCTGGGTGAATATTCACCTTCCACATTCCGTAACCCG-3'          |
| attB1                   | Forward | 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3'                       |
| attB2                   | Reverse | 5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3'                       |

**Tabla 5. Oligonucleótidos utilizados para clonar los ORFs del genoma de Bam35 en el vector donador pDONR/Zeo.** El número de ORF está anotado. La ' indica los oligonucleótidos necesarios para generar la variante sin el dominio de transmembrana del ORF indicado. El rango de aminoácidos (aa) eliminado está indicado.

Con estos oligonucleótidos y con el genoma de Bam35 como molde, se realizó una primera ronda de PCR con la DNAP Kapa siguiendo las recomendaciones del fabricante. Tras comprobar los productos de PCR en un gel al 0,7% de agarosa, se realizó una segunda ronda con el *primer forward* attB1 y el *reverse* attB2 (Tabla 5) generando los sitios attB1 y attB2 completos.

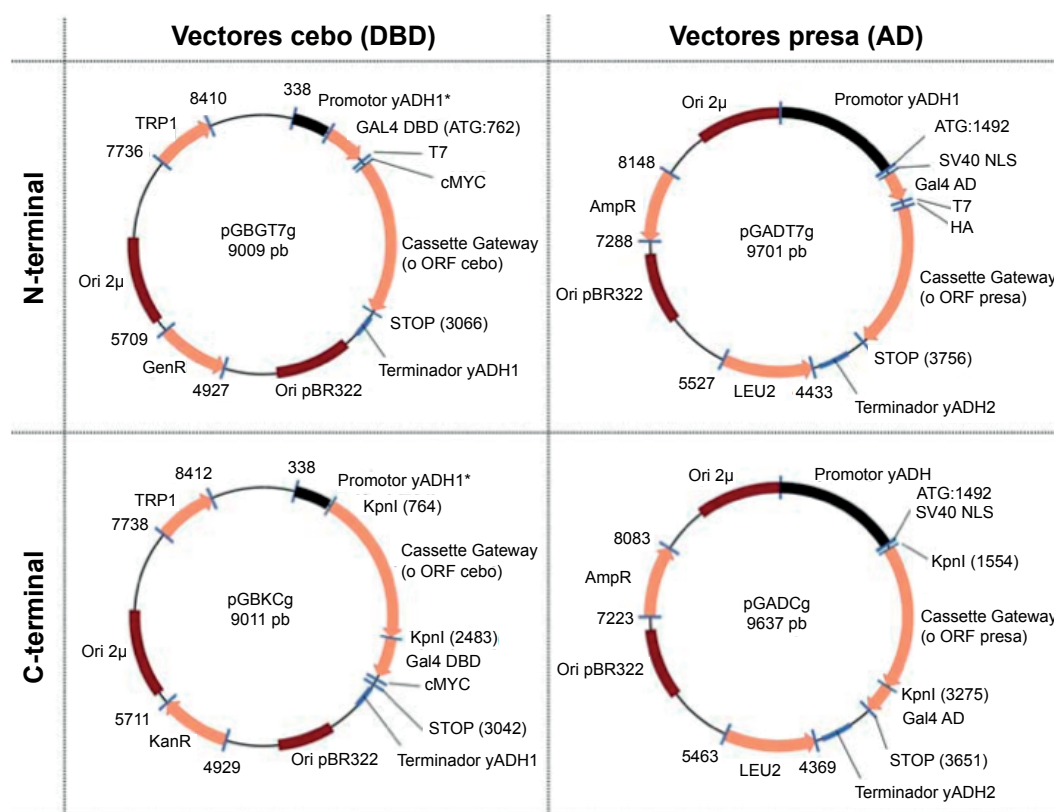
Los productos de PCR flanqueados por los sitios attB1 y attB2 fueron clonados en el vector pDONR/Zeo mediante el kit *BP Clonase™ II Enzyme Mix* (Invitrogen) siguiendo el protocolo de la casa comercial. Los productos obtenidos de la recombinación se emplearon para transformar *E. coli* DH5α y las bacterias se plaquearon en medio sólido de LB con 50 µg/mL de zeocina. Para comprobar la presencia del inserto en



las colonias, se hizo PCR de una colonia de cada uno de los clonajes utilizando los *primers* de la primera ronda de PCR. Las colonias que tuvieron un producto del tamaño esperado, se seleccionaron para aislar el plásmido y se secuenció su inserto con el *primer* universal M13 *forward*.

### 9.2. Obtención de los vectores de expresión en levaduras

Todos los ORFs de Bam35 clonados en el vector pDONR/Zeo fueron subclonados en los vectores de expresión del sistema de doble híbrido de levaduras mediante el kit *LR Clonase™ II Enzyme Mix* (Invitrogen). Los vectores utilizados fueron pGADCg (resistente a ampicilina, Amp<sup>R</sup>), pGADT7g (Amp<sup>R</sup>), pGBKCg (kanamicina resistente, Kan<sup>R</sup>) y pGBGT7g (gentamicina resistente, Gen<sup>R</sup>) (**Figura 12**) (Stellberger *et al.*, 2010). De esta manera generamos 4 colecciones de plásmidos con todos los ORFs del genoma de Bam35. Los productos resultantes se utilizaron para transformar *E. coli* DH5α y las bacterias fueron plaqueadas en medio LB sólido con el antibiótico correspondiente. Una vez aislados los vectores de nuevo, se comprobó la presencia del inserto mediante PCR con los *primers* T7 *forward* (5'-TAATACGACTCACTATAGGG-3') y el attB2 *reverse* para los clonajes en pGADT7g y pGBGT7g, y con los *primers* attB1 *forward* y attB2 *reverse* para pGADCg y pGBKCg. Además, 4 de los vectores de cada una de las colecciones fueron elegidos al azar para volverlos a secuenciar.



**Figura 12.** Esquema de los vectores utilizados en el sistema de doble híbrido en levaduras. Los vectores pGBGT7g y pGADT7g fusionan el DBD y AD de Gal4, respectivamente, en el N-terminal de la proteína del fago expresada. Los vectores pGBKCg y pGADCg fusionan el DBD y AD de Gal4, respectivamente, al C-terminal de las proteínas del fago expresada. Figura recopilada de Stellberger *et al.* (2010).

### 9.3. Sistema de doble híbrido en levaduras

Los vectores denominados cebos, pGBKCg y pGBGT7g, se emplearon para transformar la cepa haploide AH109 de *S. cerevisiae* (genotipo en **Tabla 6**) y se seleccionaron en medio sólido de levaduras sin triptófano y los vectores presa, pGADCg y pGADT7g, se emplearon para transformar la cepa Y187 (ver genotipo en **Tabla 6**) y se seleccionaron en medio sólido sin leucina, siguiendo el protocolo descrito por Mehla *et al.* (2015b).

| Cepas haploides <i>S. cerevisiae</i> | Genotipo  |
|--------------------------------------|---|
| <b>AH109 (para cebos)</b>            | MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1 <sub>UAS</sub> GAL1 <sub>TATA</sub> -HIS3, GAL2 <sub>UAS</sub> -GAL2 <sub>TATA</sub> -ADE2, URA3::MEL1 <sub>UAS</sub> -MEL1 <sub>TATA</sub> -lacZ |
| <b>Y187 (para presas)</b>            | MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met, gal80Δ, URA3::GAL1 <sub>UAS</sub> -GAL1 <sub>TATA</sub> -lacZ   |

**Tabla 6. Genotipo de las cepas de *S. cerevisiae* utilizadas para realizar el sistema de doble híbrido en levaduras.**

Posteriormente, se aparearon las levaduras para obtener levaduras diploides con los pares de vectores: pGBKCg-pGADCg, pGBKCg-pGADT7g, pGBGT7g-pGADCg y pGBGT7g-pGADT7g y se seleccionaron en medio sólido de levaduras sin leucina ni triptófano. Estas combinaciones permiten analizar la interacción entre las proteínas de fusión con los dominios del factor de transcripción Gal4 en el C- o en el N-terminal de las proteínas de Bam35 (CC, CN, NC, NN). El gen reportero utilizado fue el HIS3 (ver apartado 4, página 21) que sintetiza una imidazol-glicerol fosfato deshidratasa que es imprescindible en la biosíntesis de la histidina. Por ello, el análisis de interacción se realizó en medio selectivo sin leucina, triptófano ni histidina. De esta manera la levadura únicamente es capaz de crecer cuando las dos proteínas de Bam35 expresadas interaccionan y activan la expresión del gen HIS3. Para evitar falsos positivos se añadió un inhibidor competitivo del producto del gen HIS3 denominado 3-amino-1,2,4-triazol (3AT) cuya concentración necesaria fue establecida previamente mediante el análisis de la auto-activación de las proteínas cebo. Finalmente, el mapa de interacción entre las proteínas de Bam35 se realizó utilizando el programa *Cytoscape* (Shannon *et al.*, 2003).



## RESULTADOS

4





## Capítulo 1.

# CARACTERIZACIÓN BIOQUÍMICA DE LA DNA POLIMERASA DE BAM35



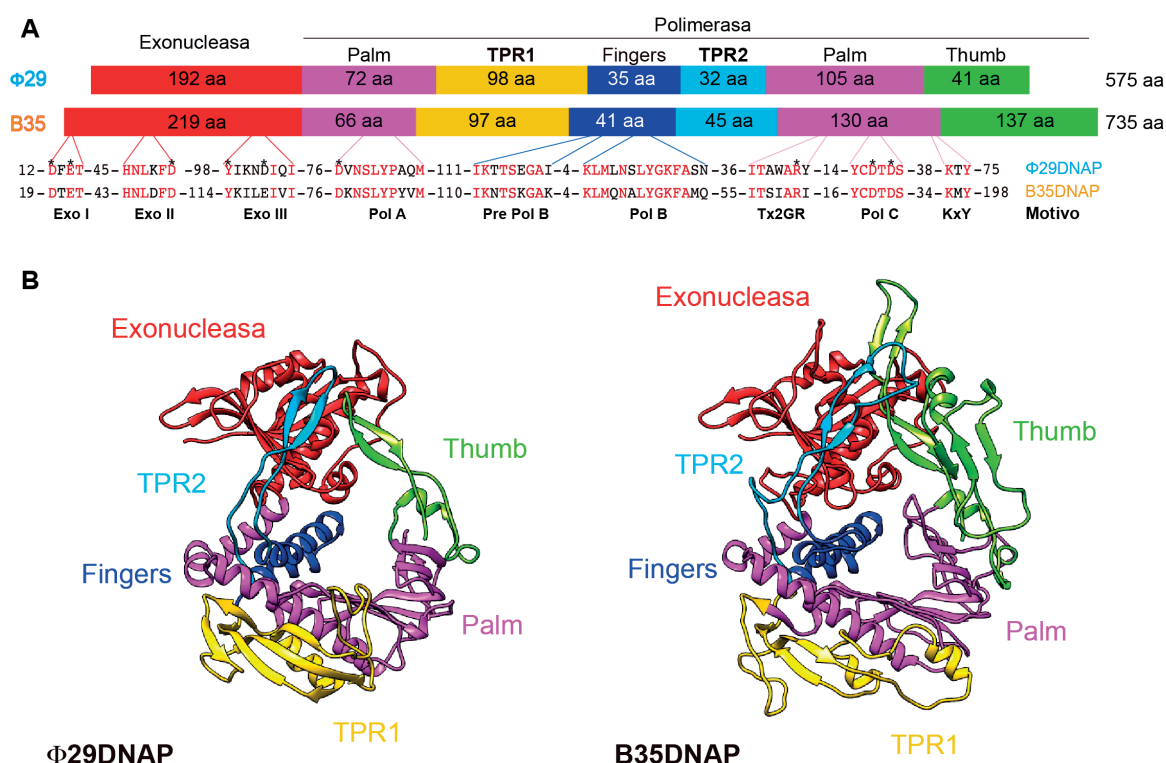
Como se ha comentado en la Introducción, la B35DNAP está codificada por el gen 5 del genoma de Bam35 y pertenece a la familia B de DNAPs replicativas.

El objetivo de este capítulo fue realizar una caracterización bioquímica y funcional inicial de la B35DNAP centrándonos en el estudio de sus actividades de polimerización y exonucleasa, en la determinación de la procesividad durante la replicación y de la capacidad intrínseca para desplazar la hebra del DNA que no es utilizada como molde. Por último, se estudió su fidelidad durante la replicación de DNA, a partir del producto de amplificación isotérmica de DNA plasmídico.

## 1. Comparación de las secuencias aminoácidicas de la B35DNAP y la $\Phi$ 29DNAP

Al realizar un *Blast* entre las secuencias aminoácidicas de la B35DNAP (735 aminoácidos) y de la  $\Phi$ 29DNAP (575 aminoácidos) se observa que comparten un 44% de similitud y un 29% de identidad. Asimismo, la organización modular de ambas DNAPs está dividida en un dominio de exonucleasa 3'-5', localizado en la región N-terminal, y un dominio de polimerización 5'-3' en la región C-terminal (**Figura 13A**) (Bernad *et al.*, 1989, Bernad *et al.*, 1990). En el dominio exonucleasa se pueden distinguir los motivos Exo I, Exo II y Exo III conservados evolutivamente tanto en DNAPs procariotas como en eucariotas (Bernad *et al.*, 1989). El dominio de polimerización presenta motivos conservados entre las DNAPs de la familia B, como por ejemplo: A (Dx<sub>2</sub>SLYP), pre B, B (Kx<sub>3</sub>NSxYG), Tx<sub>2</sub>GR, C (YxDTDS) y el KxY (revisado en Salas y de Vega (2016)). Además, este dominio está subdividido en los subdominios *palm*, *fingers*, *thumb*, y los subdominios característicos de polimerasas que inician la replicación con proteína terminal, TPR1 y TPR2 (Blasco *et al.*, 1990, Blanco *et al.*, 1991, Dufour *et al.*, 2000). Sin embargo, en el caso de la B35DNAP, el subdominio *thumb* es 96 aminoácidos más largo que el de la  $\Phi$ 29DNAP.

Para obtener una visión estructural general de la B35DNAP, realizamos un modelado con el servidor *I-TASSER* utilizando como modelo la estructura tridimensional de la  $\Phi$ 29DNAP (Kamtekar *et al.*, 2004, Yang y Zhang, 2015). En este modelo, la disposición general de los subdominios es muy similar a la de la  $\Phi$ 29DNAP. Sin embargo, observamos que principalmente las zonas menos conservadas, *thumb* y TPR2, muestran una menor similitud en su estructura terciaria (**Figura 13B**). Durante el transcurso de esta tesis, el Dr. Miguel García Díaz intentó obtener la estructura cristalográfica de la B35DNAP sin obtener resultados positivos hasta el momento.



**Figura 13. A) Representación esquemática de las secuencias peptídicas de la B35DNAP y la  $\Phi$ 29DNAP.** Los dominios y subdominios conservados están representados como rectángulos de colores y su longitud en aminoácidos está indicada dentro. Los residuos idénticos están resaltados en rojo y los residuos implicados en la unión de DNA y en la catálisis están indicados con asteriscos. **B) Estructura tridimensional de la B35DNAP y la  $\Phi$ 29DNAP.** Los diferentes subdominios están indicados en los mismos colores que en el panel A. La estructura de la  $\Phi$ 29DNAP fue obtenida por Kamtekar *et al.* (2004), la de la B35DNAP es un modelo realizado por el servidor I-TASSER.

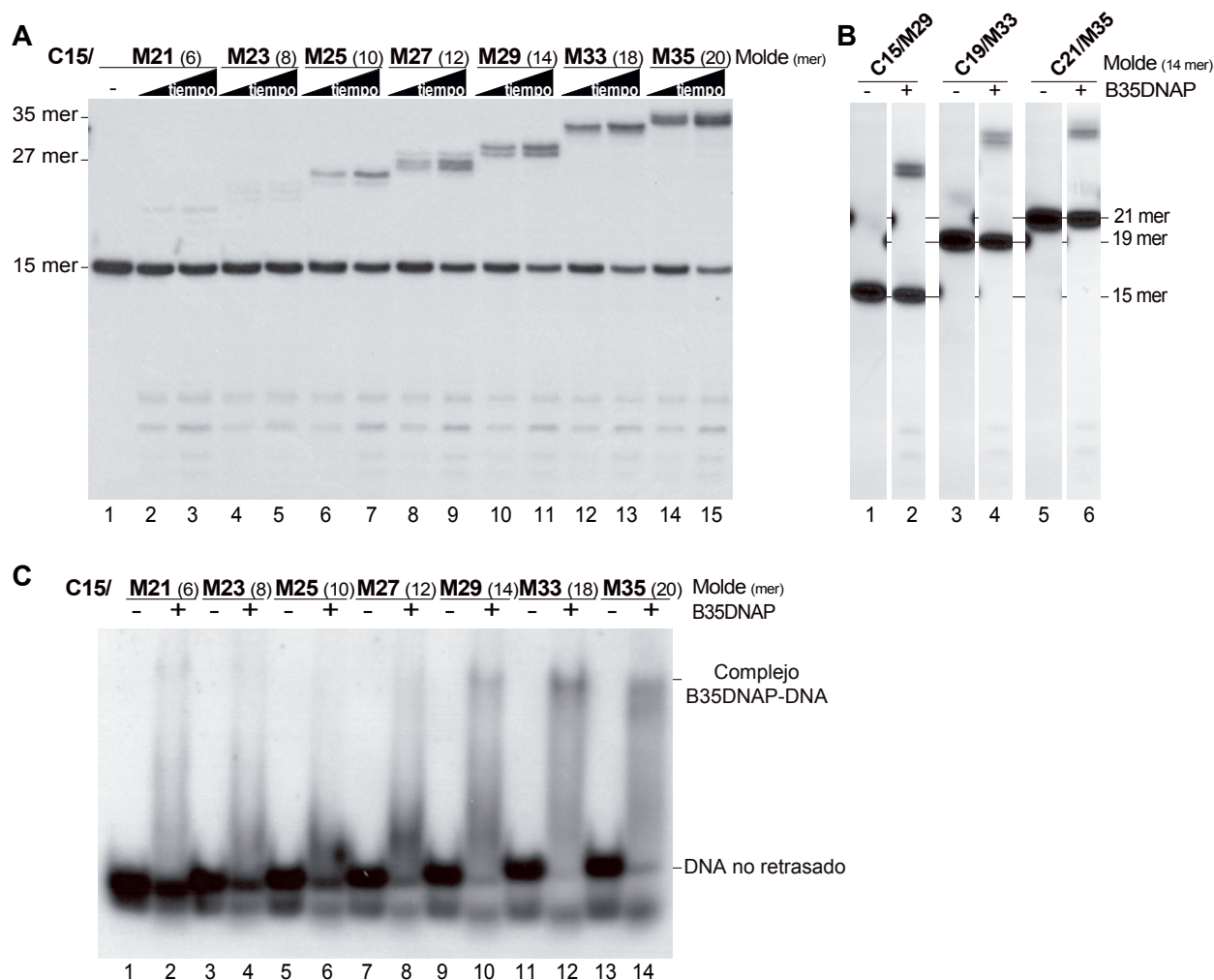


## 2. Optimización del sustrato para ensayos de replicación de oligonucleótidos por la B35DNAP

Con el objetivo de determinar si la B35DNAP recombinante era funcional, analizamos su capacidad de polimerización mediante la extensión de varios sustratos de oligonucleótidos formados por un *primer* (C15) hibridado con distintos moldes (M21, M23, M25, M27, M29, M33GTT y M35). El ensayo se realizó a alta concentración de dNTPs (1  $\mu$ M) durante 10 y 30 min (Materiales y Métodos 8.1, página 49). El resultado (**Figura 14A**) mostró que el producto de replicación correspondiente al tamaño completo del molde se obtenía cuando el número de nucleótidos en el extremo 5' protuberante del molde era de 10 o superior (carriles 6-15), mientras que cuando era inferior prácticamente no se observaba elongación del *primer* ni siquiera a 30 min de incubación (carriles 2-5). Además, en todos los casos, excepto en el control negativo (carril 1) se observaron productos de degradación debidos, posiblemente, a la actividad exonucleasa de la B35DNAP. Para determinar si el tamaño del dúplex de DNA formado por el *primer*/molde también afectaba a la síntesis de la B35DNAP, se realizó el mismo ensayo con tres *primers* diferentes hibridados con distintos moldes con el fin de que todos los moldes tuviesen 14 nucleótidos en su extremo 5' protuberante (C15/M29, C19/M33GTT y C21/M35).



La **Figura 14B** muestra que, a diferencia del tamaño del molde, el tamaño del dúplex *primer*/molde no afecta a la polimerización.

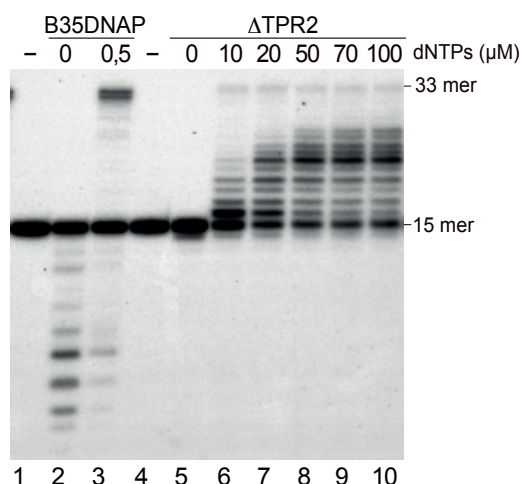


**Figura 14. A) Ensayo de replicación de oligonucleótidos de DNA con el molde de distintos tamaños por la B35DNAP.** El ensayo se llevó a cabo con los sustratos *primer*/molde indicados a 1  $\mu$ M de dNTPs durante 10 y 30 min a 37 °C. El tamaño de 15-mer corresponde al *primer* y los tamaños mayores al *primer* elongado. **B) Ensayo de replicación de oligonucleótidos de DNA con el dúplex *primer*/molde de diferente tamaño.** El ensayo se realizó como el del panel A pero únicamente a 10 min. Los tamaños 15, 19 y 21-mer corresponden a los *primers*. **C) Retraso en gel con los mismos sustratos del panel A.** Las incubaciones se llevaron a cabo con 2 nM de sustrato, 20 mM sulfato amónico, 10 mM de  $MgCl_2$  y en presencia o ausencia de 5 nM de B35DNAP.

Una posible causa de la escasa polimerización mostrada por la B35DNAP con moldes pequeños podría ser el requerimiento de un molde más grande para el establecimiento de una interacción estable entre la enzima y el molde. Para comprobar esta hipótesis, llevamos a cabo un experimento de retraso en gel con los mismos sustratos que el ensayo del panel A (Materiales y Métodos 8.2, página 50). El resultado mostrado por el panel C de la **Figura 14** indicó que la B35DNAP establece una interacción estable con el sustrato *primer*/molde cuando el molde posee más de 12 nucleótidos en su extremo 5' protuberante (carriles 10, 12 y 14 frente a 2, 4, 6 y 8). En conjunto, estos resultados indican que la B35DNAP es una enzima funcional con actividad de polimerización 5'-3'. Además, sugieren que dicha DNAP requiere de un molde de al menos 10

nucleótidos para establecer una unión estable y poder elongar eficazmente el *primer*. Por ello, el sustrato *primer*/molde más utilizado en esta tesis fue el C15/M33.

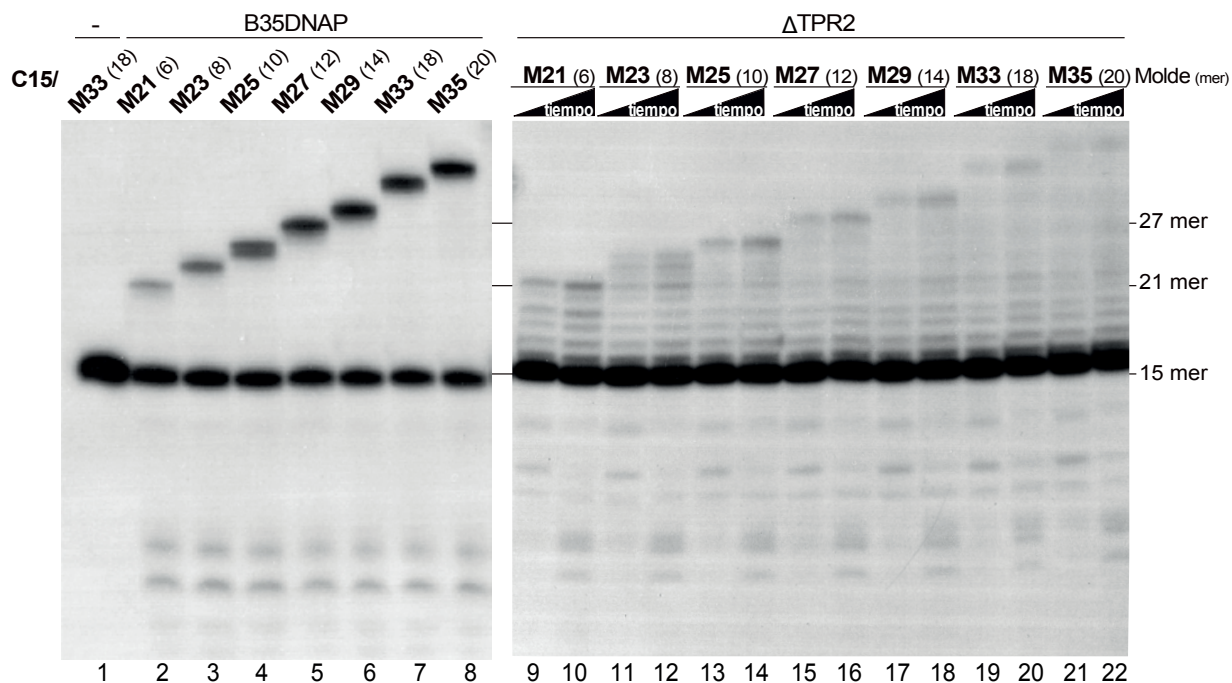
Con el objetivo de intentar explicar el requerimiento de un molde largo para establecer una unión estable, analizamos el modelado de la estructura cristalográfica de la B35DNAP (**Figura 13B**). Puesto que el TPR2 de la  $\Phi$ 29DNAP interacciona con la cadena molde del DNA (Kamtekar *et al.*, 2004, Berman *et al.*, 2007), nos preguntamos si el subdominio TPR2 de la B35DNAP (mayor que el de la  $\Phi$ 29DNAP) estaría implicado en el requerimiento de un molde largo. Para estudiar esta hipótesis, generamos una variante de delección del subdominio TPR2 de la B35DNAP ( $\Delta$ TPR2, Materiales y Métodos 4, página 41). Inicialmente, analizamos la funcionalidad de esta variante mediante el estudio de su capacidad para degradar y extender el *primer* del dúplex de DNA C15/M13 (**Figura 15**). Al igual que su equivalente en la  $\Phi$ 29DNAP (Rodríguez *et al.*, 2005), la variante  $\Delta$ TPR2 de la B35DNAP mostró una polimerización distributiva (**Figura 15**, carriles 6-10) y una disminución en su actividad exonucleasa (carril 5). Posteriormente, analizamos su capacidad de extensión del *primer* C15 hibridado a los mismos moldes utilizados con la B35DNAP *wild type* en la **Figura 14A**.



**Figura 15. Actividades exonucleasa y polimerasa de la variante  $\Delta$ TPR2 de la B35DNAP.** El ensayo se realizó con 200 nM del  $\Delta$ TPR2 con el sustrato C15/M33GTT a las concentraciones indicadas de dNTPs y se incubó 30 min a 37 °C (Materiales y Métodos 8.1, página 49). En los carriles 2 y 3 se muestra las actividades exonucleasa y polimerasa de la B35DNAP *wild type* (10 nM, 10 min de incubación) como control. Las posiciones de los oligonucleótidos de 15 y 33-mer están indicados a la derecha.

En este caso utilizamos una concentración mayor de nucleótido (100  $\mu$ M) por lo que analizamos de nuevo la replicación producida por la B35DNAP *wild type* (**Figura 16**). Al incrementar la concentración de nucleótido observamos que la B35DNAP *wild type* fue capaz de replicar los oligonucleótidos molde con los extremos 3' protuberantes más pequeños (6 y 8-mer, **Figura 16**, carriles 2-3), sin embargo, la eficiencia de replicación seguía siendo inferior que con el molde a partir de 12 nucleótidos protuberantes (carriles 2, 3 frente a 5-8). Por el contrario, el  $\Delta$ TPR2 replicó con la misma eficiencia los moldes con el extremo 3' protuberante más pequeños (6, 8 y 10-mer, **Figura 16**, carriles 9-14) que el de 12 nucleótidos protuberantes (carriles 15 y

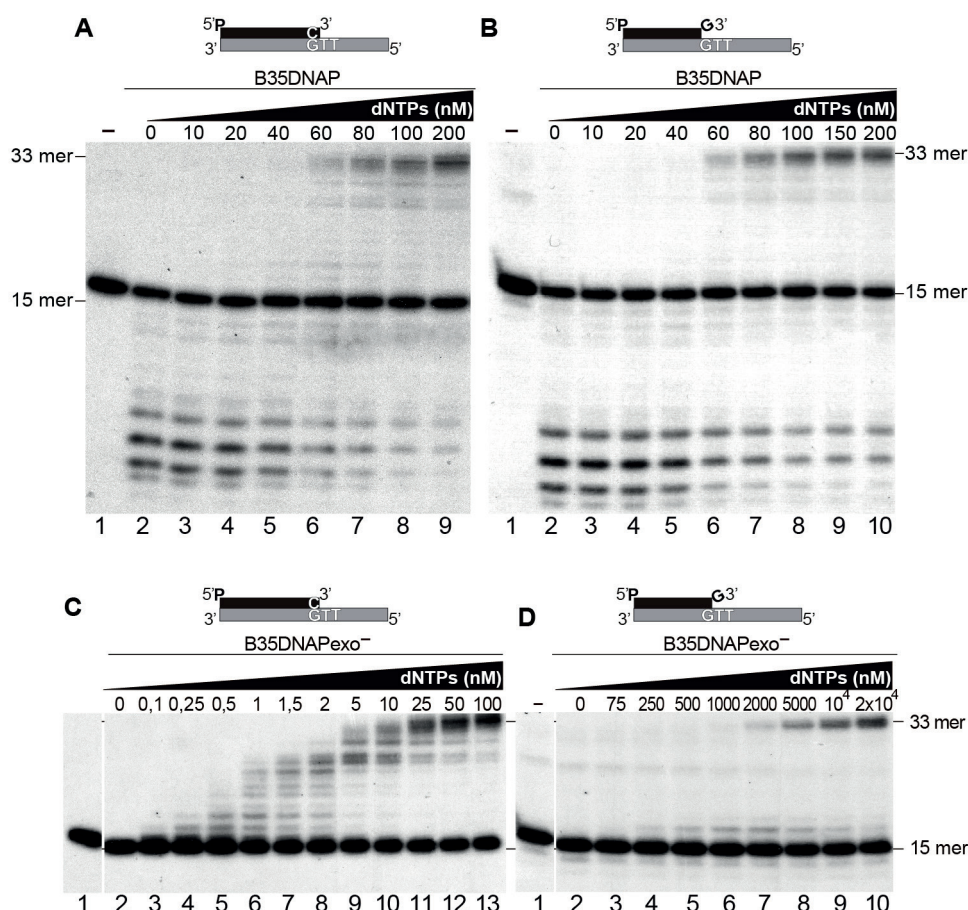
16). La pequeña menor eficacia en la replicación de moldes aún más largos (14, 18 y 20-mer, carriles 17-22) probablemente se deba a la pérdida de procesividad mostrada por esta variante. Estos resultados sugieren una implicación del subdominio TPR2 en la unión al DNA molde necesario para la replicación del mismo.



**Figura 16. Ensayo de replicación de oligonucleótidos de DNA con el molde de distintos tamaños por la variante  $\Delta$ TPR2 de la B35DNAP.** El ensayo se llevó a cabo con los sustratos *primer/molde* indicados a 100  $\mu$ M de dNTPs. El panel izquierdo corresponde al control positivo llevado a cabo con 10 nM de la B35DNAP *wild type* durante 10 min de incubación. El panel de la derecha muestra el resultado obtenido con 200 nM de la variante  $\Delta$ TPR2 en las mismas condiciones pero durante una incubación de 10 y 30 min a 37 °C. El tamaño de 15-mer corresponde al *primer* y los tamaños mayores al *primer* elongado.

### 3. Las actividades de polimerización y exonucleasa de la B35DNAP están coordinadas

Una vez establecido el sustrato óptimo evaluamos la actividad de degradación 3'-5' y analizamos el equilibrio entre la actividad exonucleasa y de polimerización de la B35DNAP, mediante el estudio de la degradación y extensión del sustrato C15/M33GTT en el cual el *primer* es completamente complementario al molde y del sustrato C15G/M33GTT en el cual el nucleótido del extremo 3' del *primer* está desapareado (Materiales y Métodos 8.1, página 49). Con ambos sustratos, en ausencia de dNTPs sólo se observaron productos de degradación (**Figura 17**, geles A y B, carril 2) mientras que, al incrementar su concentración, la actividad exonucleasa fue reemplazada progresivamente por la actividad de polimerización requiriéndose la misma cantidad de dNTPs (60 nM, carril 6 en geles A y B) para replicar el tamaño completo de ambos sustratos. Este resultado indica que la actividad correctora de errores de la B35DNAP es capaz de eliminar el nucleótido desapareado del extremo 3' del *primer* para posteriormente extenderlo.

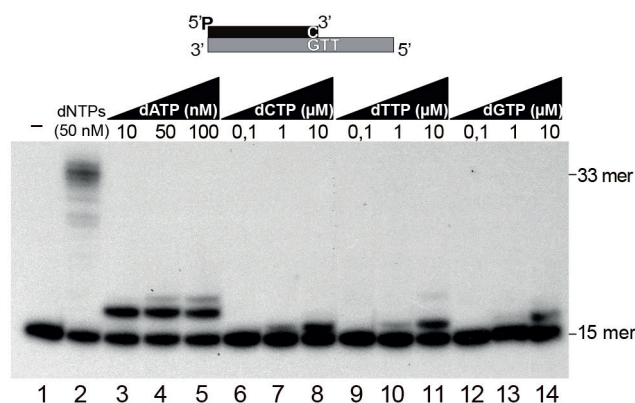


**Figura 17. Polimerización y degradación de oligonucleótidos de DNA por la B35DNAP.** Los geles desnaturalizantes de acrilamida de los paneles A y C muestran los resultados obtenidos con el *primer* (C15) complementario al molde (M33GTT) con 10 nM de B35DNAP o B35DNAPexo-, respectivamente. Los geles de los paneles B y D muestran los resultados con el *primer* (C15G) que tiene un nucleótido en el extremo 3' desapareado con respecto al molde (M33GTT) y con 10 nM de B35DNAP o B35DNAPexo-, respectivamente. La concentración de dNTPs y una representación del *primer*/molde utilizados están indicados. La posición correspondiente al *primer* de 15-mer y al producto obtenido de 33-mer están marcados. La incubación se realizó durante 10 min a 37 °C.

Aprovechando que los residuos catalíticos de la B35DNAP están conservados, realizamos una variante de la enzima deficiente en la actividad exonucleasa (B35DNAPexo-) mediante el cambio del ácido aspártico de la posición 19 y del ácido glutámico 21 a alanina (**Figura 13**; Materiales y Métodos 4, página 41). Esta nueva variante de la B35DNAP se utilizó con el fin de confirmar el papel de la actividad exonucleasa en la extensión del *primer* C15G mediante la comparación, de nuevo, de las concentraciones de dNTPs necesarias para elongar el *primer* con su extremo 3' apareado (C15) frente al desapareado (C15G) (**Figura 17**, paneles C y D). En este ensayo se observó que la extensión del *primer* C15G requirió alrededor de 1.000 veces más concentración de dNTPs que el C15 (carril 9 del gel C frente al 8 del D). La variante sin actividad exonucleasa también nos permitió determinar la especificidad de inserción de nucleótido. Así, la inserción de un nucleótido incorrecto precisa, aproximadamente, una concentración 100 veces más alta que en el caso del nucleótido correcto (**Figura 18**). Por lo tanto, tal y como se espera para una DNAP replicativa,



el sitio activo de polimerización de la B35DNAP es altamente específico para la inserción del nucleótido que aparea correctamente con el molde y, además, en caso de errores, la B35DNAP es capaz de eliminarlo mediante su actividad correctora de pruebas 3'-5'.



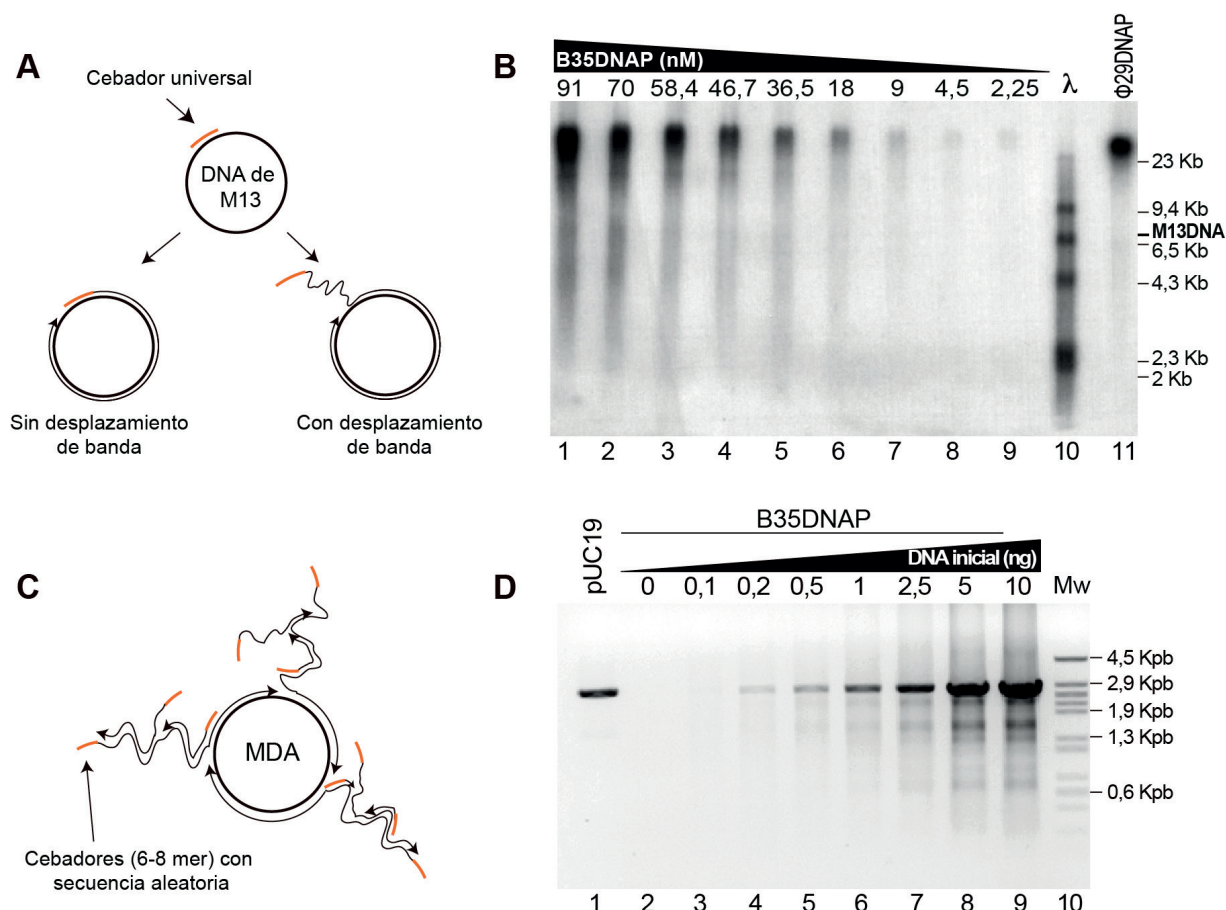
**Figura 18. Fidelidad en la inserción de nucleótido por la B35DNAP.** El gel desnaturalizante de acrilamida muestra la preferencia de inserción de nucleótido por la B35DNAPexo- usando el *primer* C15 que es totalmente complementario al molde M33GTT (esquema en la parte superior del gel). Las reacciones se incubaron durante 10 min a 37 °C a la concentración de cada uno de los dNTPs indicados.

#### 4. La B35DNAP sintetiza el DNA de manera procesiva acoplada al desplazamiento de banda

Con el objetivo de analizar la capacidad de la B35DNAP para acoplar la polimerización al desplazamiento de la hebra de DNA que no actúa como molde, llevamos a cabo un ensayo de replicación por círculo rodante con el genoma de cadena sencilla del bacteriófago M13 hibridado al *primer* universal (**Figura 19A**) (Materiales y Métodos 8.3, página 50). Como se muestra en la **Figura 19B**, la B35DNAP fue capaz de replicar el DNA de M13 dando lugar a un producto del mismo tamaño que el obtenido con la  $\Phi$ 29DNAP, siendo, en ambos casos superior a la longitud del genoma de M13 (7,25 Kb) (carriles 1-9 frente al 11). Este resultado indica que la B35DNAP es capaz de abrir la doble hélice de DNA, desplazando la hebra complementaria a la hebra molde durante la síntesis de DNA. Además, la replicación de la B35DNAP es procesiva puesto que el tamaño máximo del producto replicado permaneció invariable, incluso cuando la concentración de enzima se diluyó 40 veces (carril 1 frente al 9).

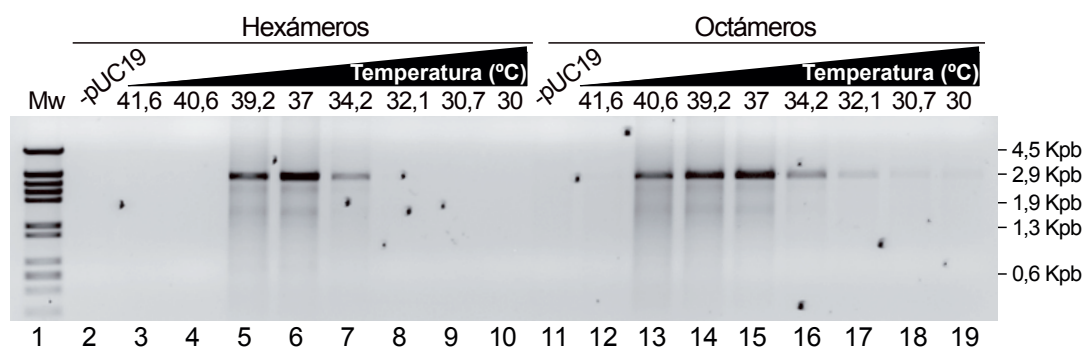
Gracias a la capacidad de desplazamiento de banda y la alta procesividad de la  $\Phi$ 29DNAP, esta DNAP es capaz de llevar a cabo una amplificación isotérmica de DNA por desplazamiento múltiple (MDA) utilizando como *primers* oligonucleótidos de 6-8-mer con secuencia al azar (**Figura 19C**) (Nelson *et al.*, 2002). Puesto que la B35DNAP también muestra estas características, quisimos determinar si era capaz de amplificar el plásmido pUC19 mediante dicho mecanismo (Materiales y Métodos 8.4, página 51). Como muestra la **Figura 19D**, la B35DNAP amplificó el plásmido utilizando *primers* de 6-mer con secuencia al azar dando

lugar a un producto de dsDNA, ya que aproximadamente el 80% del producto amplificado fue linearizado al digerir con la endonucleasa EcoRI. Además, la mayor parte del producto de amplificación correspondió con repeticiones en tándem del pUC19 inicial puesto que las bandas observadas en la electroforesis fueron del mismo tamaño que el pUC19 original (2,7 Kpb, carril 1). En estas condiciones la B35DNAP fue capaz de amplificar una cantidad inicial de DNA de 1-2,5 ng (carriles 6 y 7), cantidad similar a la utilizada de manera rutinaria en la MDA de plásmidos de DNA por la  $\Phi$ 29DNAP (Nelson *et al.*, 2002, Nelson, 2014).



**Figura 19. Polimerización procesiva acoplada a desplazamiento de banda durante la replicación de DNA circular por la B35DNAP.** A) Esquema de los posibles productos obtenidos en la replicación del DNA de M13 (modificado de Blanco *et al.* (1989)). El tamaño del producto obtenido será igual que el tamaño unidad del DNA molde si la DNAP no es capaz de desplazar banda. Si por el contrario, la DNAP acopla su replicación al desplazamiento de banda, se obtendrá un producto mayor que el tamaño unidad del DNA molde. B) Ensayo de replicación del DNA circular de M13. Las concentraciones de B35DNAP utilizadas están indicadas; en el caso de la  $\Phi$ 29DNAP se utilizó 50 nM. La longitud del DNA sintetizado se analizó en gel alcalino al 0,7% de agarosa utilizando como marcador el DNA de  $\lambda$  digerido con HindIII y marcado radiactivamente (carril 10) (Materiales y Métodos 8.3, página 50). C) Esquema de la MDA de DNA. Este ensayo utiliza un exceso de *primers* pequeños (6-8-mer) con secuencias aleatorias lo que les permite hibridar en diferentes zonas del DNA. Los *primers* son elongados por una DNAP con capacidad de desplazamiento de banda generando nuevas cadenas de ssDNA que son susceptibles de hibridar con más *primers*. Así se genera una amplificación altamente ramificada. D) MDA del plásmido de DNA pUC19 por la B35DNAP. El ensayo se realizó a 50 nM de B35DNAP con las cantidades de plásmido inicial indicados (Materiales y Métodos 8.4, 51). Los productos se linearizaron digiriendo con EcoRI antes de la electroforesis en gel al 0,7% de agarosa. Como control de tamaño se cargaron 100 ng del plásmido pUC19 linearizado (carril 1) y DNA de  $\Phi$ 29 digerido con HindIII (carril 10).

Finalmente, estudiamos la termoestabilidad de la B35DNAP durante la MDA. Puesto que la temperatura de fusión ( $T_m$ ) de los *primers* se incrementa con su longitud, decidimos utilizar, además de los *primers* de 6-mer (hexámeros), *primers* de 8-mer (octámeros). De esta manera, disminuimos la posibilidad de que la ausencia de amplificación se debiese a la imposibilidad del *primer* para hibridar con el DNA molde como consecuencia de las altas temperaturas. Con ambos tipos de *primers*, la temperatura óptima de amplificación de la B35DNAP fue de 37 °C (**Figura 20**, carriles 6 y 15). Sin embargo, mientras que con hexámeros el producto empezaba a disminuir a 39,2 °C y no era detectable a 40,6 °C, con octámeros, la amplificación a 39,2 °C fue prácticamente la misma que a 37 °C, alcanzando a los 40,6 °C, la temperatura máxima tolerada (carriles 4, 5 frente a 13, 14). Estos resultados sugieren que la ausencia de actividad a 40,6 °C con hexámeros podría deberse a la mayor dificultad de los *primers* para hibridar con el plásmido pUC19 a dicha temperatura. Por otro lado, con respecto a las temperaturas más bajas, observamos que a 34,2 °C la amplificación fue la misma con ambos tipos de *primers* (carriles 7 y 16) mientras que con octámeros, a diferencia de con los hexámeros, también se detectaba algo de amplificación a 32,1 °C. Esto podría deberse a una mayor capacidad de los *primers* más largos para hibridar con el DNA a cualquier temperatura.



**Figura 20. Termoestabilidad de la B35DNAP en la MDA.** El ensayo se llevó a cabo con 1 ng de pUC19 inicial, 50 nM de B35DNAP y con *primers* de 6 o 8-mer (carriles 2-10 o 11-19, respectivamente) durante 16 horas a las temperaturas indicadas. Como control negativo se realizó el mismo ensayo en ausencia de pUC19 (carriles 2 y 11). Como marcador de tamaño se cargaron 100 ng del DNA de  $\Phi$ 29 digerido con HindIII (carril 1).



## 5. La B35DNAP amplifica el DNA con alta fidelidad siendo su principal error la inserción de bases en secuencias homopoliméricas

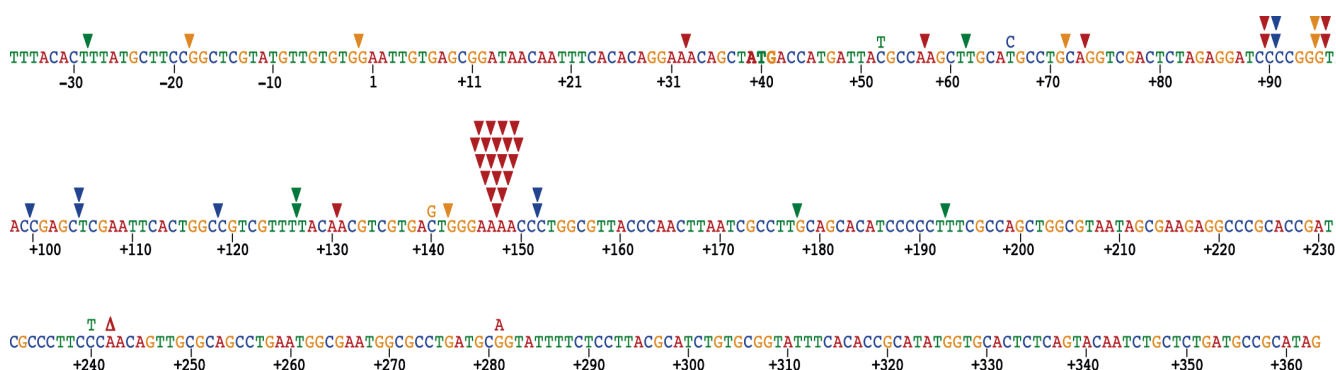
Aprovechando la presencia del gen *lacZ $\alpha$*  en el plásmido pUC19, analizamos la fidelidad de la B35DNAP en la amplificación de DNA. Para ello, el producto linearizado procedente de la MDA fue re-circularizado y utilizado para transformar *E. coli* XL-1 Blue con el objetivo de identificar mutaciones en el gen *lacZ $\alpha$*  mediante el fenotipo de las colonias. De esta manera, en presencia de IPTG se induce la expresión del gen *lacZ $\alpha$*  dando lugar a la  $\beta$ -galactosidasa que, si es funcional, hidroliza el sustrato x-Gal dando lugar a galactosa y 5-bromo-4-cloro-3-hidroxindol, producto que a su vez se oxida generando un compuesto insoluble de

color azul llamado 5,5'-dibromo-4,4'-dicloro-índigo (Cepko *et al.*, 1998). Así, mediante el análisis de la relación entre el número de colonias blancas (con mutaciones) y las totales se puede determinar la tasa de error durante la replicación de las DNAPs (Materiales y Métodos 8.5, página 51). Tal y como se muestra en la **Tabla 7**, la media de la tasa de error de la B35DNAP fue de  $5,1 \pm 1,7 \times 10^{-6}$ , fidelidad similar a la  $\Phi 29$ DNAP y otras polimerasas comerciales utilizadas en sistemas de amplificación de DNA (Cline *et al.*, 1996, Nelson *et al.*, 2002, McInerney *et al.*, 2014).

| Experimento    | Colonias totales | f                            | f <sub>0</sub>               | d              | Tasa de error                |
|----------------|------------------|------------------------------|------------------------------|----------------|------------------------------|
| 1              | 1.171            | $1,8 \times 10^{-2}$         | $5,7 \times 10^{-3}$         | 11,2           | $3,2 \times 10^{-6}$         |
| 2              | 10.605           | $2,7 \times 10^{-2}$         | $4,6 \times 10^{-3}$         | 9,9            | $6,5 \times 10^{-6}$         |
| 3              | 10.682           | $2,1 \times 10^{-2}$         | $2,2 \times 10^{-3}$         | 10             | $5,6 \times 10^{-6}$         |
| Media $\pm$ DE |                  | $1,9 \pm 0,2 \times 10^{-2}$ | $4,2 \pm 1,6 \times 10^{-3}$ | $10,4 \pm 0,9$ | $5,1 \pm 1,7 \times 10^{-6}$ |

**Tabla 7. Determinación del rango de error de la B35DNAP durante la MDA.** La frecuencia de mutación (f) fue determinada dividiendo el número total de colonias blancas entre el número total de colonias. La f<sub>0</sub> representa la frecuencia de mutación del fondo dado por el pUC19 purificado de la bacteria, digerido y religado en las mismas condiciones. La duplicación del molde (d) fue calculado mediante la ecuación  $\log_2$  (ng de pUC19 finales/ng de pUC19 iniciales). La tasa de error fue calculada con la ecuación  $(f-f_0)/pb \times d$ , donde pb es el número de nucleótidos del gen lacZ (342 pb) (Materiales y Métodos 8.5, página 51). En la tabla se indican los datos de tres experimentos independientes y la media  $\pm$  desviación estándar (DE).

Para determinar el error más frecuente cometido por la B35DNAP, secuenciamos el gen lacZ $\alpha$  de 50 de las colonias blancas. La **Figura 21** muestra que el cambio más común fue la inserción de nucleótidos en las zonas homopoliméricas de la secuencia (con una frecuencia de  $2,7 \times 10^{-3}$  entre los clones mutantes) acumulándose la mayor parte de las mutaciones en dos regiones concretas del gen lacZ $\alpha$ : 5'-GGGAAAACCC-3' y 5'-CCCCGGG-3'. De todas las colonias secuenciadas sólo una presentó una deleción y sólo se detectaron cinco substituciones, cuatro de ellas fueron transiciones (C:G $\Rightarrow$ T:A) y la otra fue una transversión (C:G $\Rightarrow$ G:C).



**Figura 21. Esquema de las mutaciones producidas por la B35DNAP en la MDA.** En rojo se representa la adenina (A), en azul la citosina (C), en verde la timina (T) y en naranja la guanina (G). La inserción de nucleótidos está indicada con un triángulo invertido ( $\blacktriangledown$ ), la deleción con  $\Delta$  y la sustitución con la base mutada, manteniendo el mismo código de colores. La posición 1 corresponde al primer nucleótido transcrito del gen lacZ $\alpha$ .



Con estos resultados se puede concluir que la B35DNAP es capaz de amplificar DNA circular mediante la MDA con una frecuencia de error muy baja, aunque la presencia de secuencias homopoliméricas podrían ser una fuente de acumulación de mutaciones por inserción, lo que implicaría cambios en el marco de lectura.





## Capítulo 2.

# ESTUDIO DE LAS PRIMERAS ETAPAS DE REPLICACIÓN DEL DNA DE BAM35



Como se menciona en la introducción, la presencia en la B35DNAP de los subdominios TPR1 y TPR2 característicos de las DNAPs que inician la replicación utilizando como *primer* una TP, la existencia de repeticiones terminales invertidas (ITRs) en los extremos del genoma del fago y la posibilidad de que su DNA esté unido de manera covalente a algún tipo de proteína (Blasco *et al.*, 1990, Ravantti *et al.*, 2003), sugieren fuertemente que la replicación llevada a cabo por la B35DNAP está iniciada por una TP. Esta proteína podría ser codificada por el ORF4 del genoma viral (Sozhamannan *et al.*, 2008, Redrejo-Rodríguez *et al.*, 2012, Jalasvuori *et al.*, 2013). El principal objetivo de este capítulo fue comprobar este mecanismo de replicación mediante la caracterización de la función de la proteína codificada por ese ORF, además de determinar el residuo de la TP utilizado por la B35DNAP para añadir el primer nucleótido de la replicación y de establecer la posición del nucleótido del extremo 3' del genoma utilizado como molde para la iniciación. Los resultados obtenidos nos permiten sugerir un mecanismo de *jumping-back* de un nucleótido como mecanismo de recuperación de la información genética de los extremos del genoma.

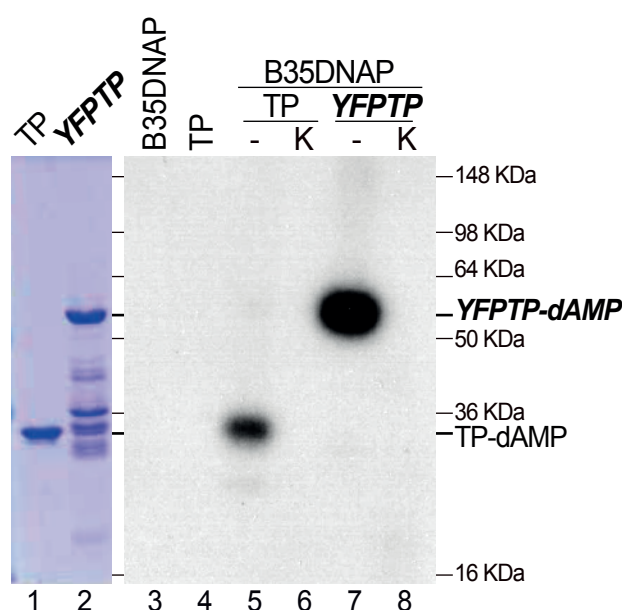


### 1. El ORF4 del genoma de Bam35 codifica una TP funcional

Tal y como se explica en la introducción, el paso de iniciación de la replicación del DNA mediada por TP conlleva la unión del primer deoxinucleótido a un grupo OH de dicha proteína. Esta reacción está dirigida por la complementariedad con un nucleótido de la hebra molde del DNA y es llevada a cabo por la DNAP. Además, generalmente, cuando se añade manganeso en ensayos de iniciación *in vitro*, se puede observar una deoxinucleotidilación de la TP independiente de molde (Blanco *et al.*, 1992, Caldentey *et al.*, 1992).

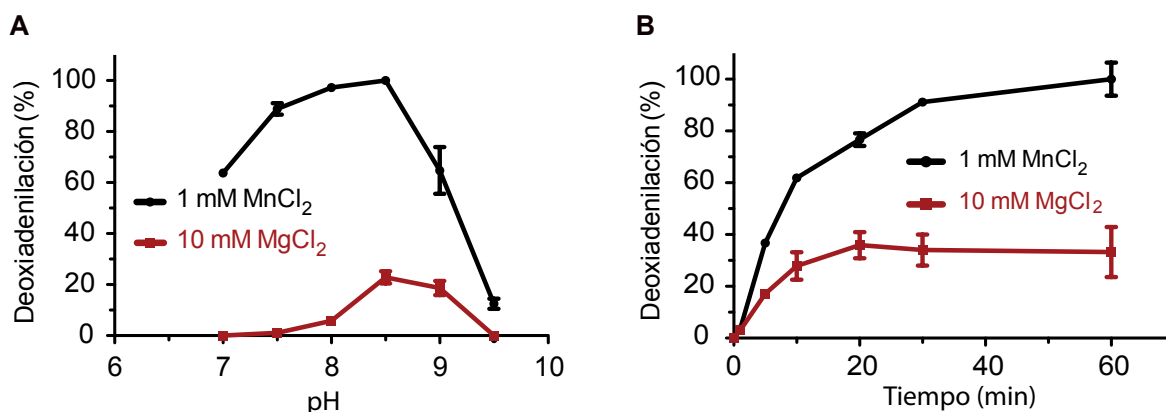
Con el objetivo de confirmar que la proteína codificada por el ORF4 del genoma de Bam35 es una TP funcional, se expresó y purificó la proteína recombinante (Materiales y Métodos 5, página 43). Posteriormente, se analizó si la hipotética TP podía ser deoxiadenilada por la B35DNAP a través de un ensayo denominado “iniciación sin molde”, utilizando manganeso como cofactor (Materiales y Métodos 7.1, página 46). Esta reacción dio lugar a un producto marcado radiativamente únicamente cuando la TP (o la TP fusionada a YFP, YFPTP) y la B35DNAP estaban presentes (**Figura 22**, carriles 3 y 4 frente a 5 y 7). El producto obtenido cuando la TP fue añadida tuvo la misma movilidad electroforética, alrededor de 30 KDa, que la TP detectada mediante tinción con azul de Coomassie (carriles 1 y 5). Cuando la TP fue sustituida por la YFPTP la banda migró menos (carriles 2 y 7), ya que la proteína de fusión YFPTP tiene un tamaño mayor que la TP (59 KDa frente a 29 KDa). En la tinción con azul de Coomassie de la YFPTP (carril 2) se observaron bandas de menor tamaño que podrían corresponder a la proteína de fusión degradada. Sin embargo, estas proteínas no contendrían la TP funcional puesto que no aparecen marcadas radiativamente al realizar el ensayo de deoxiadenilación. Tanto en el caso de la TP como con la YFPTP, la banda radiactiva desapareció al tratar con proteinasa K (Materiales y Métodos 7.1, página 46) confirmando que el producto corresponde a

una proteína marcada con  $^{32}\text{P}$ -dAMP (carriles 6 y 8).



**Figura 22. Ensayo funcional de la TP de Bam35.** Gel 12% SDS-PAGE. Los carriles 1 y 2 muestran 1  $\mu\text{g}$  de las proteínas purificadas TP e YFPTP, respectivamente. Los carriles 3-8 corresponden a un ensayo de deoxiadenucleación en condiciones estándar con manganeso (Materiales y Métodos 7.1, página 46). Los carriles 3 y 4 son los controles del ensayo sin TP y sin B35DNAP, respectivamente. Los carriles 5 y 7 muestran las bandas correspondientes a la TP y a la YFPTP marcadas radiactivamente. Los carriles 6 y 8 son el resultado del tratamiento con proteinasa K de los productos obtenidos en los carriles 5 y 7.

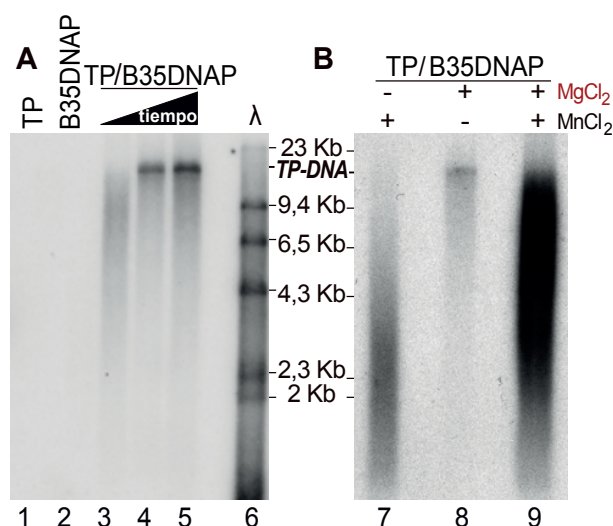
Para estudiar con más profundidad la reacción de deoxiadenucleación de la TP por la B35DNAP independiente de molde, se analizó el efecto de los metales divalentes manganeso y magnesio y, debido al alto punto isoeléctrico de la TP, se analizó el efecto del pH. La **Figura 23A** muestra que, con ambos metales, la reacción es más eficiente a pH ligeramente básico siendo este efecto mucho más pronunciado en el caso del magnesio. De hecho, con magnesio no fue posible detectar ningún producto de deoxiadenucleación hasta que el pH fue de 8 y se obtuvo la mayor eficiencia a pH 8,5. En el caso del manganeso apenas existieron diferencias entre los pHs 7,5-8,5 aunque el óptimo, al igual que con magnesio, se obtuvo a pH 8,5. Debido a estos resultados, y para facilitar su estudio, la mayoría de los ensayos de iniciación de esta tesis se realizaron con manganeso a pH 7,5 y, cuando se indica, el papel del magnesio fue estudiado, utilizando, en este caso, un pH de 8,5. Con estas condiciones, se quiso determinar si a lo largo del tiempo la reacción de iniciación con magnesio era capaz de alcanzar la misma eficiencia que con manganeso, para lo cual se realizó una cinética de iniciación sin molde. Como se puede observar en la **Figura 23B**, el porcentaje de producto obtenido con magnesio fue entre 2 y 3 veces menor que con manganeso indicando que la iniciación sin molde se produce preferentemente con este último metal.



**Figura 23. Estudio del efecto del pH y de la dependencia de metal en la reacción de deoxiadenilación de la TP de Bam35 independiente de molde.** A) Iniciación sin molde a distintos pHs con manganeso (MnCl<sub>2</sub>) o magnesio (MgCl<sub>2</sub>) como cofactor. Las reacciones se llevaron a cabo en las condiciones estándar excepto por los diferentes pHs los cuales fueron obtenidos utilizando Tris-HCl para los pHs de 7 a 8,5 y CHES-NaOH para los pHs 9 y 9,5 (Materiales y Métodos 7.1, página 46). B) Cinética de la reacción de iniciación independiente de molde con MnCl<sub>2</sub> a pH 7,5 y MgCl<sub>2</sub> a pH 8,5. En ambas gráficas se representa la media del porcentaje de deoxiadenilación relativo a la reacción más eficiente (100%) de tres experimentos independientes, además de su correspondiente DE.

A continuación, nos preguntamos si, además de ser sustrato para la reacción de deoxiadenilación independiente de molde, la TP de Bam35 podría funcionar como *primer* durante la replicación del genoma de Bam35 *in vitro*. Así, realizamos una incubación del TP-DNA de Bam35 (Materiales y Métodos 6, página 45) con la TP y la B35DNAP en presencia de los cuatro dNTPs y con magnesio como cofactor (Materiales y Métodos 7.6, página 48). Mediante un gel de agarosa alcalino al 0,7% se observó que a la hora de incubación no existía una banda clara de replicación pero sí algunos productos más pequeños (**Figura 24A** carril 3). Sin embargo, a las 2 y 4 horas se apreció una banda de alrededor de 15 Kb que corresponde con el tamaño del genoma de Bam35 (carriles 4 y 5). En ausencia de la B35DNAP o de la TP no se observó ningún tipo de producto de replicación (carril 1 y 2, respectivamente). Este resultado indica que la TP puede ser empleada como *primer* por la B35DNAP para la replicación completa del genoma de Bam35, de acuerdo con las hipótesis previas.

Además, puesto que la deoxiadenilación de la TP en ausencia de molde ocurre preferentemente con manganeso, quisimos analizar si al utilizar ese metal divalente sólo o junto a magnesio se incrementaba la eficiencia de la replicación del TP-DNA con respecto a la obtenida con magnesio. No obstante, la presencia de magnesio fue necesaria para obtener un producto de replicación del tamaño del genoma de Bam35 mientras que la presencia de manganeso, tanto sólo como junto al magnesio, produjo un incremento en la señal radiactiva pero disminuyó la capacidad para alcanzar el tamaño completo del TP-DNA (**Figura 24B** carriles 7-9). Este resultado podría deberse a una pérdida de la procesividad de la B35DNAP cuando utiliza el manganeso como cofactor tal y como se ha demostrado previamente en otros sistemas similares como PRD1 (Caldentey *et al.*, 1992).



**Figura 24. Replicación del genoma de Bam35 iniciada con TP.** Electroforesis en gel de agarosa alcalino al 0,7%. Las reacciones se llevaron a cabo con 11 nM de B35DNAP, 133 nM de TP y 100 ng de TP-DNA de Bam35 (ver Materiales y Métodos 7.6, página 48). A) Replicación con 10 mM de MgCl<sub>2</sub>. Las reacciones se incubaron durante 1, 2 y 4 horas (carriles 3-5). Los carriles 1 y 2, corresponden a incubaciones durante 4 horas de los controles sin B35DNAP y sin TP, respectivamente. B) Efecto del metal sobre la replicación de TP-DNA. Las reacciones se activaron con 10 mM de MgCl<sub>2</sub> y/o 1 mM de MnCl<sub>2</sub> como se indica y se incubaron durante 2 horas. Como indicador de tamaño se utilizó el DNA de λ digerido con HindIII marcado radiactivamente. El tamaño esperado para el TP-DNA de Bam35 está indicado.

Con estos resultados podemos concluir que la deoxiadenuilación de la TP y la replicación completa del TP-DNA de Bam35 *in vitro* son posibles con la presencia de únicamente dos proteínas, la TP y la B35DNAP y, por lo tanto, que el ORF4 del genoma de Bam35 codifica la TP viral, esencial para la replicación del genoma iniciada con proteína terminal.

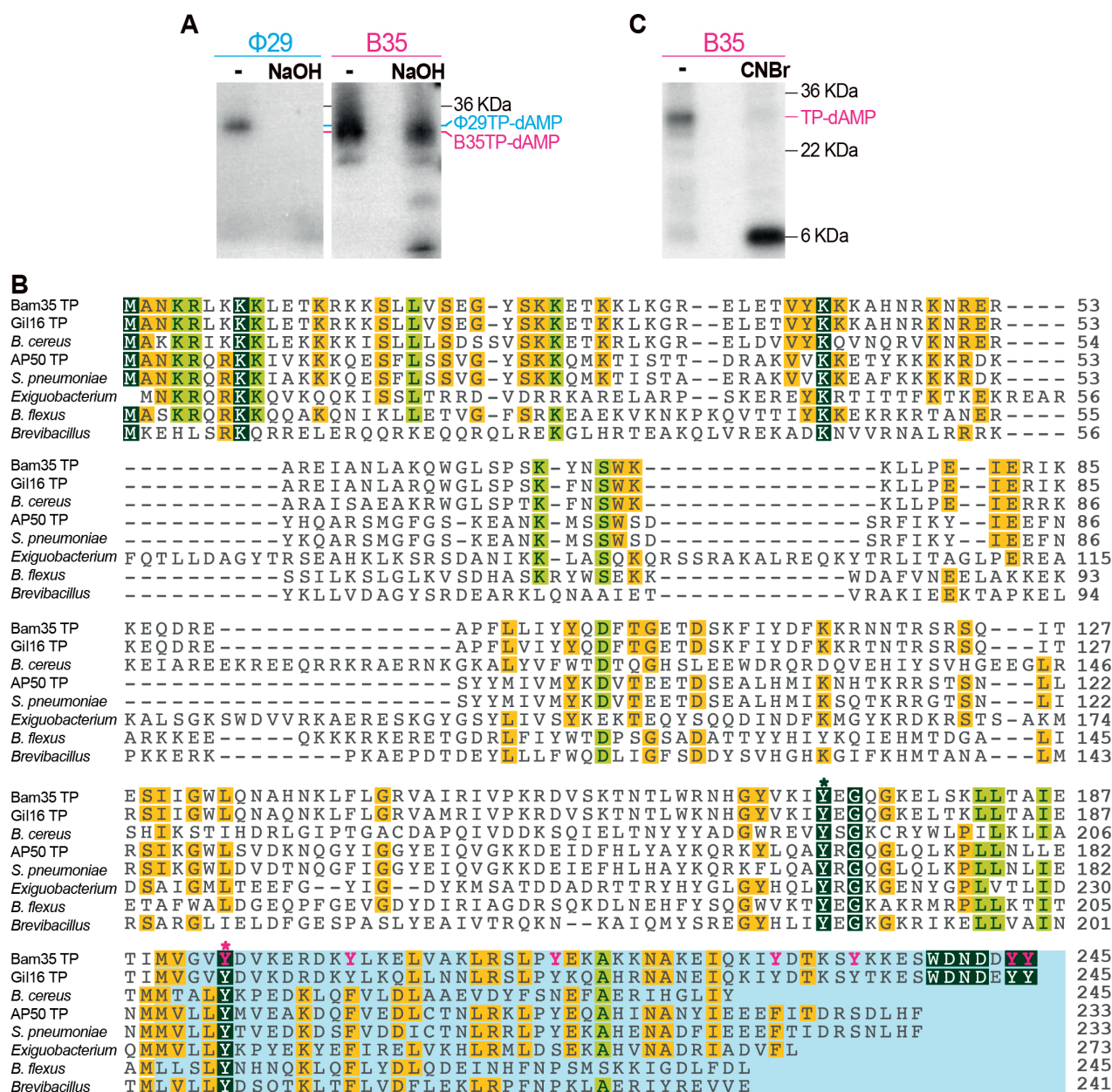


## 2. La tirosina conservada 194 de la TP de Bam35 es el residuo iniciador de la replicación del DNA

La reacción de iniciación en la replicación de DNA mediada por TP consiste en la adición, catalizada por la DNAP, de un deoxinucleótido al grupo hidroxilo proporcionado por una serina, treonina o tirosina de la TP (Salas, 1991). Con el fin de determinar cual de estos tres aminoácidos es utilizado en la TP de Bam35, una TP previamente deoxiadenuilada (TP-dAMP) por la B35DNAP en ausencia de molde fue sometida a tratamiento alcalino con calor (Materiales y Métodos 7.2, página 47). Este tratamiento rompe el enlace fosfoéster entre los residuos serina o treonina y el deoxinucleótido mientras que, si está unido a una tirosina el enlace es resistente (Rothberg *et al.*, 1978, Salas, 1991). Como muestra la **Figura 25A**, el complejo de Φ29 TP-dAMP, en el cual el residuo iniciador de la TP es la serina 232 (Hermoso *et al.*, 1985), fue sensible, tal y como se preveía, a las condiciones alcalinas; sin embargo, el complejo TP-dAMP de Bam35 fue resistente. Este resultado sugiere fuertemente que el residuo iniciador de la TP de Bam35 es una tirosina.

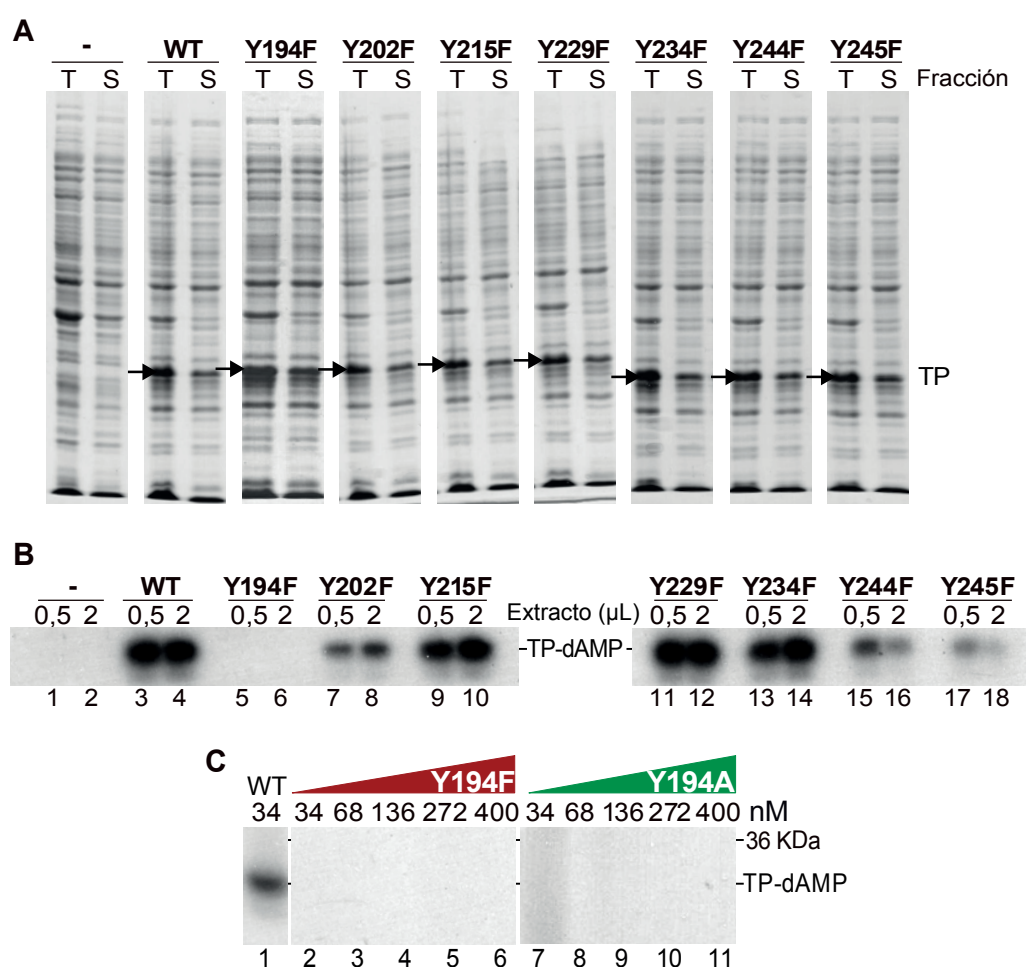
Al analizar la secuencia aminoacídica de la TP de Bam35 (**Figura 25B**) se observa que contiene 15 tirosinas aunque sólo dos, las de las posiciones 172 y 194, están totalmente conservadas entre secuencias similares. Además, sólo presenta dos metioninas, una en la posición 1 y otra en la 190. Esta característica fue aprovechada para delimitar la localización de la tirosina iniciadora de la TP de Bam35 mediante tratamiento con CNBr, que produce la escisión de las cadenas polipeptídicas por la parte C-terminal de las metioninas. Por tanto, en el caso de Bam35, esperaríamos únicamente dos fragmentos, uno de 22,4 KDa (que comprendería





**Figura 25. Mapeo del residuo iniciador de la TP de Bam35.** A) Determinación de la naturaleza del residuo iniciador de la TP de Bam35 mediante tratamiento alcalino. La reacción de iniciación con molde con la DNAP y la TP de Φ29 se utilizó como control y se realizó de forma paralela a la de Bam35. Posteriormente, ambas reacciones fueron incubadas en presencia o ausencia de 100 mM NaOH durante 6 min a 95°C. Finalmente, se neutralizaron y se analizaron en gel 12% SDS-PAGE y autorradiografía. B) Alineamiento múltiple de secuencias de la TP de Bam35 y otras TP's relacionadas. Las secuencias mostradas corresponden a hipotéticas TP's de tectivirus representativos que infectan a bacterias Gram-positivas: Bam35 (NCBI ID NP\_943750.1), Gil16 (YP\_224102.1), AP50 (YP\_002302516.1), además de otras secuencias ortólogas obtenidas mediante BLAST desde la base de datos de proteínas NR del NCBI ([www.ncbi.nlm.nih.gov/refseq/](http://www.ncbi.nlm.nih.gov/refseq/)). Estas secuencias están anotadas como posibles TP's de las bacterias *B. cereus* (WP\_001085581.1), *Streptococcus pneumoniae* (WP\_050224775.1), *Exiguobacterium antarcticum* (WP\_026829749.1), *Bacillus flexus* (WP\_025907183.1) y *Brevibacillus* sp. CF112 (WP\_007784052.1). Las secuencias fueron alineadas con el algoritmo MUSCLE implementado en el programa Geneious R8 (Kearse *et al.*, 2012). El fragmento C-terminal correspondiente a la fragmentación de las proteínas con el CNBr está sombreado en azul y las tirosinas de la TP de Bam35 presentes en este fragmento están marcadas en rosa. Las tirosinas conservadas 172 y 194 están marcadas con asteriscos encima de la secuencia. C) Mapeo del residuo iniciador de la TP de Bam35. El complejo TP-dAMP se incubó con 1,2 mM de CNBr y 200 mM de HCl durante 20 horas a temperatura ambiente. Finalmente, las muestras fueron neutralizadas y analizadas en gel 18% SDS-PAGE. Ver Materiales y Métodos 7.2 de la página 47, para más detalles.

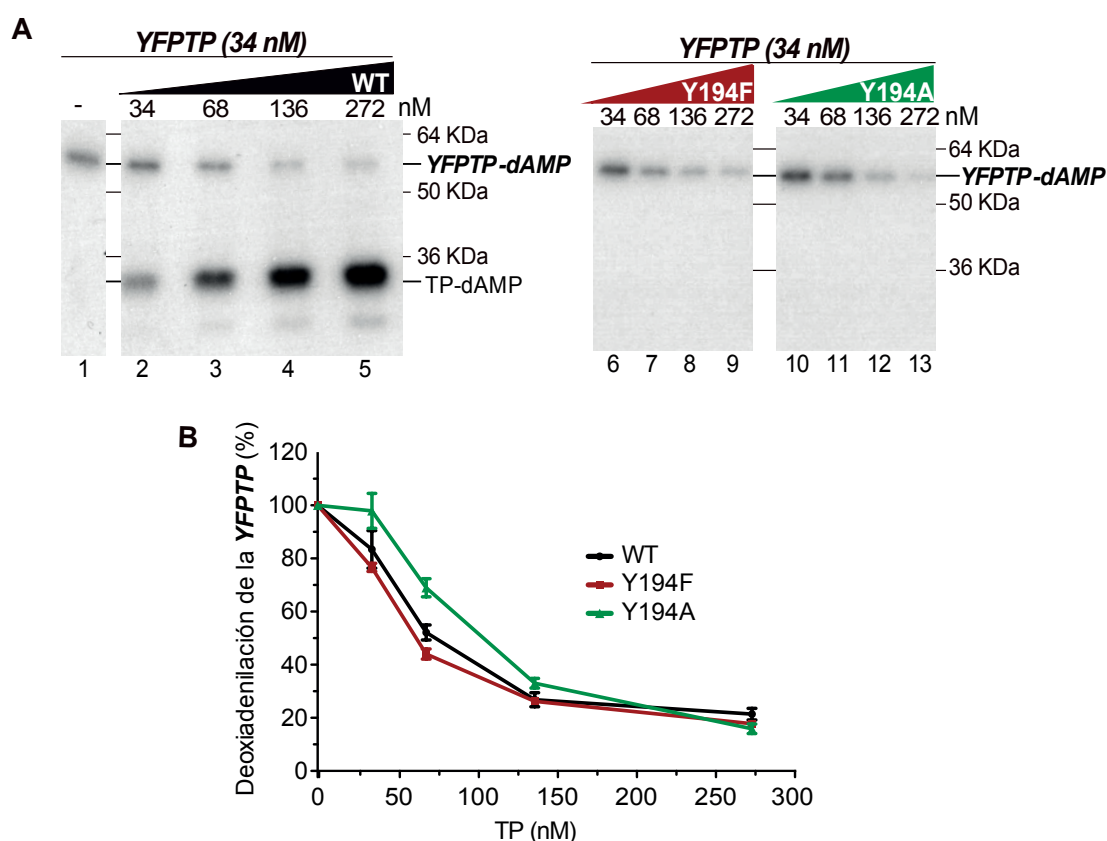
del aminoácido 2 al 190) y el otro de 6,2 KDa (con los aminoácidos 191-245). De esta manera podemos determinar en cuál de los dos fragmentos de la TP está unido el dAMP marcado radiactivamente. El tratamiento del complejo TP-dAMP con CNBr (Materiales y Métodos 7.2, página 47) mostró que sólo el fragmento de 6,2 KDa mantuvo el marcaje radiactivo (**Figura 25C**) indicando que la tirosina iniciadora se encontraba en la parte C-terminal de la proteína y más concretamente entre los residuos 191-245. En este fragmento hay un total de 7 tirosinas, por lo tanto, para determinar cuál de ellas era utilizada por la B35DNAP para iniciar la replicación, realizamos mutagénesis dirigida para convertir todas esas tirosinas en fenilalaninas (marcadas en rosa en la **Figura 25B**). Puesto que la reacción de iniciación es muy específica y requiere la interacción entre la TP y la DNAP, se puede realizar dicha reacción con los extractos bacterianos que expresan las diferentes variantes en las tirosinas de la TP. Previamente se verificó que las proteínas expresadas eran solubles y que



**Figura 26. Identificación del residuo iniciador de la TP de Bam35.** A) Expresión de las distintas variantes en las tirosinas de la TP de Bam35 en extractos de *E.coli* BL21(DE3). Gel 12% SDS-PAGE teñido con azul de Coomassie de las fracciones totales (T) y solubles (S) de dichos extractos. Las bandas que corresponden a las TPs están indicadas. B) Identificación de la tirosina 194 como residuo iniciador mediante ensayos de deoxiadenuclilación sin molde con 0,5 o 2  $\mu$ L de extractos bacterianos de cultivos que expresan las variantes de las TPs. En el gel al 12% SDS-PAGE los carriles 1 y 2 corresponden a los extractos bacterianos transformados con el plásmido vacío (control negativo), los carriles 3 y 4 con el plásmido de expresión de la TP *wild type* (control positivo) y los carriles 5-18 con las distintas variantes de la TP de Bam35. C) Deoxiadenuclilación de la TP independiente de molde de la TP *wild type* (carril 1) y de concentraciones crecientes de las TPs Y194F y Y194A purificadas. Las reacciones se llevaron a cabo en las condiciones estándar. Ver Materiales y Métodos 7.3, página 47.

la expresión entre ellas era similar (**Figura 26A**). A continuación, se realizó el ensayo de iniciación sin molde con 0,5 y 2  $\mu\text{L}$  de los extractos bacterianos solubles (Materiales y Métodos 7.3, página 47) observándose que todos los mutantes excepto el Y194F fueron deoxiadenuclados (**Figura 26B**). Este resultado fue confirmado mediante el mismo ensayo, pero utilizando la proteína TP Y194F y la variante no conservativa Y194A purificadas no detectándose producto ni con 10 veces más concentración que la TP *wild type* (**Figura 26C**). Ambos resultados sugieren fuertemente que la tirosina 194 es el residuo iniciador de la TP de Bam35.

La ausencia de deoxiadenuclación podía ser debida a una pérdida de la funcionalidad de esas variantes de la TP y no a que el aminoácido cambiado fuese el residuo iniciador. Para descartar esta posibilidad, se comparó la capacidad de las TPs Y194F y Y194A para interaccionar con la B35DNAP frente a la capacidad de interacción funcional de la YFPTP. Así, se valoró el descenso de la deoxiadenuclación de la YFPTP producida como consecuencia de la incubación previa de la B35DNAP con cantidades limitantes de la TP *wild type* o de las variantes Y194F o Y194A. La proteína *wild type* y la variante conservativa Y194F interaccionaron con la B35DNAP de manera similar puesto que el incremento en la concentración de ambas proteínas dio



**Figura 27. Estudio de la interacción funcional entre las variantes de la TP de Bam35 Y194F e Y194A y la B35DNAP.** A) Inhibición de la deoxiadenuclación de la YFPTP por concentraciones crecientes de la TP *wild type* o de las variantes Y194F o Y194A. Las reacciones se llevaron a cabo con 1 mM  $\text{MnCl}_2$  en presencia de las TPs indicadas y, después de 2,5 min de reacción, la proteína de fusión YFPTP fue añadida para competir por la unión a la B35DNAP y las muestras se incubaron otros 2,5 min. B) Representación gráfica del efecto de las diferentes concentraciones de las variantes de las TPs sobre el porcentaje de deoxiadenuclación de la YFPTP. La media y la DE de tres experimentos independientes están indicadas. Ver Materiales y Métodos 7.4, página 48.

lugar a un patrón similar en la inhibición de la deoxiadenucleación de la YFPTP (**Figura 27A** carriles 2-5 y 6-9 y líneas negra y roja en **Figura 27B**). Por otro lado, el mutante Y194A requirió una concentración ligeramente más alta para inhibirla (**Figura 27A** carriles 10-13 y **Figura 27B** línea verde) aunque no es una diferencia significativa. Estos resultados confirman que la mutación en la tirosina 194 sólo afecta a la capacidad iniciadora de la TP lo cual indica que ese residuo actúa como anclaje para el primer nucleótido de la replicación.



### **3. La tirosina conservada 172 de la TP de Bam35 no es necesaria para la iniciación pero está implicada en la interacción con la B35DNAP**

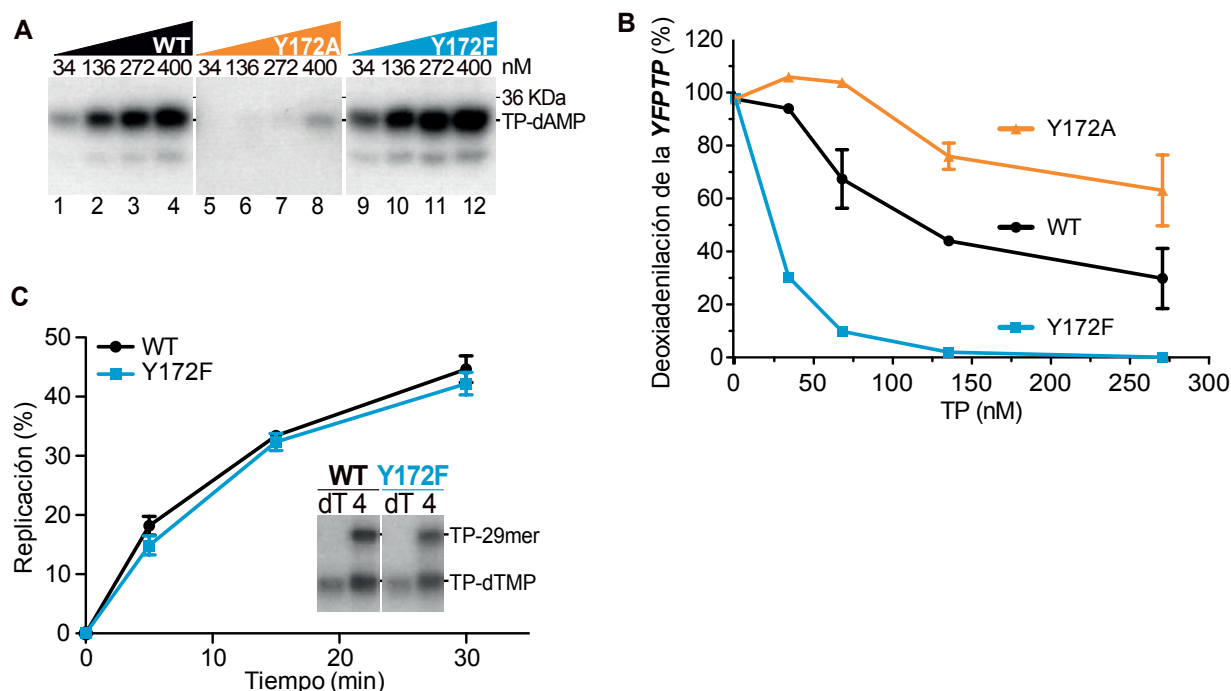
Como se menciona en el apartado anterior, las únicas tirosinas conservadas entre las secuencias de TPs relacionadas con Bam35 son la 194 y la 172 (**Figura 25B**). Por este motivo, nos preguntamos si la tirosina 172 podría servir como un residuo iniciador secundario o estar implicado en otro tipo de función. Con el objetivo de estudiar estas posibilidades, analizamos la capacidad de las variantes de la TP Y172F e Y172A para ser deoxiadenucladas por la B35DNAP. En la **Figura 28A** se observa que la TP Y172A fue casi inactiva y que la banda correspondiente a la proteína marcada sólo se detectó cuando se utilizó la concentración más alta (400 nM, carril 8). Sin embargo, la deoxiadenuclación de la variante conservativa Y172F fue algo superior a la observada con la TP *wild type* (carriles 9-12 frente a 1-4). De esta manera, podemos concluir que la tirosina 172 no está directamente involucrada en la deoxinucleotidilación de la TP.

Para estudiar en más detalle el papel del residuo 172, analizamos la interacción de las TPs Y172A e Y172F con la B35DNAP mediante el análisis de la reducción competitiva de la deoxiadenuclación de la YFPTP. En la **Figura 28B** se observa que el mutante Y172A, a las concentraciones ensayadas, no fue capaz de disminuir el marcaje de la YFPTP en más de un 20-30% lo que sugiere que esta proteína podría estar afectada en la interacción con la B35DNAP. Por el contrario, a concentraciones equimolares de la variante Y172F y la YFPTP, la variante disminuyó el producto YFPTP-dAMP mucho más que la *wild type* llegando incluso a una reducción del 70% en la deoxiadenuclación. Esto sugiere una interacción más competitiva de la TP Y172F con la B35DNAP respecto a la proteína *wild type*.

Esta mayor estabilidad en la interacción podría afectar a la disociación entre la B35DNAP y la TP producida después de las primeras etapas de la replicación y que es esencial para continuar con la polimerización. Para analizar esta posibilidad, estudiamos la iniciación dependiente de molde seguida de la replicación de un oligonucleótido de 29-mer de cadena sencilla cuya secuencia es la misma que el origen izquierdo del genoma de Bam35 (Materiales y Métodos 7.5, página 48 y oligonucleótido WT de la **Tabla 3**). La B35DNAP mostró una eficiencia similar en la replicación del oligonucleótido con ambas proteínas (TP *wild type* y su variante Y172F) sugiriendo que la mejora en la interacción mostrada por el mutante Y172F no afecta a los



pasos siguientes de la replicación (Figura 28C).



**Figura 28. Caracterización funcional del residuo conservado Y172 de la TP de Bam35.** A) Iniciación independiente de molde con concentraciones crecientes de la TP *wild type*, y sus variantes Y172A e Y172F. Las concentraciones de proteína utilizadas están indicadas. B) Análisis comparativo de la interacción funcional de la TP *wild type* y las variantes Y172A e Y172F con la B35DNAP. La gráfica representa el efecto de las distintas concentraciones de las TP estudiadas sobre el porcentaje de deoxiadenuilación de la YFPTP. La media y la DE de tres experimentos independientes están indicadas (Materiales y Métodos 7.4, página 48). C) Replicación iniciada por TP de un oligonucleótido de 29-mer de cadena sencilla cuya secuencia es la misma que la del origen izquierdo del genoma de Bam35. El ensayo se realizó a distintos tiempos con las TP *wild type* o Y172F con  $MnCl_2$  (Materiales y Métodos 7.5, página 48). En la gráfica se representa la media y la DE del porcentaje de replicación con respecto a la iniciación de tres experimentos independientes. El panel de la derecha muestra un gel representativo 15% SDS-PAGE de los productos de iniciación (con dTTP) y replicación (con los cuatro dNTPs) iniciados con la TP *wild type* o su variante Y172F después de 30 min de reacción.

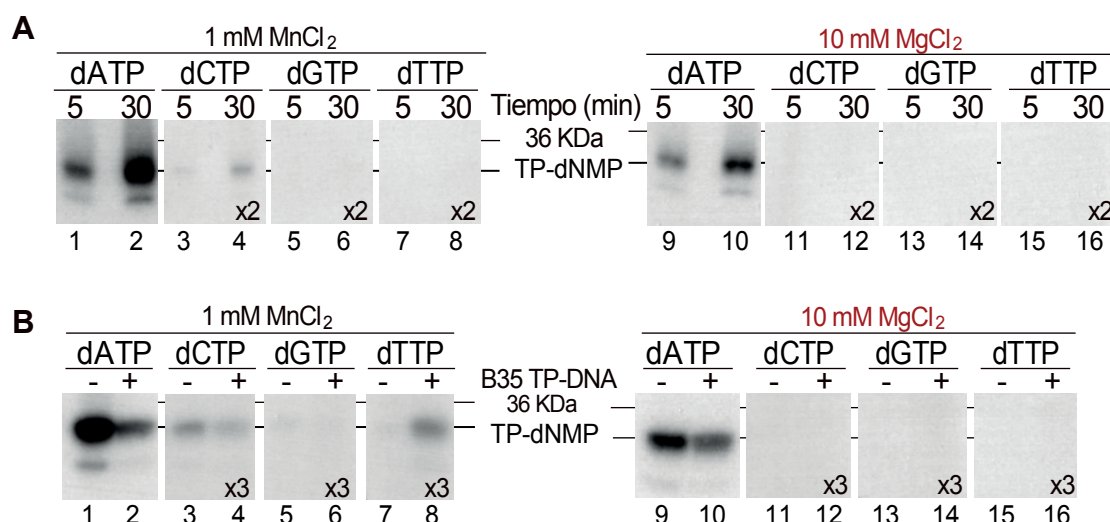
Estos resultados indican que el residuo conservado Y172 de la TP de Bam35 puede estar involucrado en la interacción con la B35DNAP y, por lo tanto, confirman que el residuo Y194 es el único aminoácido conservado que puede ser utilizado para la unión del primer nucleótido en la replicación el genoma de Bam35.

#### 4. La iniciación de la replicación del genoma de Bam35 está dirigida principalmente por el tercer nucleótido del extremo 3' de la cadena molde

Con el objetivo de estudiar los primeros pasos de la replicación del genoma de Bam35 analizamos que deoxinucleótido era unido preferentemente a la TP por la B35DNAP en ausencia y presencia de TP-DNA como molde. La deoxinucleotidilación de la TP independiente de molde fue principalmente dependiente de dATP tanto cuando se utilizó manganeso como magnesio como metal divalente (Figura 29A carriles 1-2 y

9-10) y sólo una pequeña incorporación de dCMP fue detectada con manganeso (carril 4).

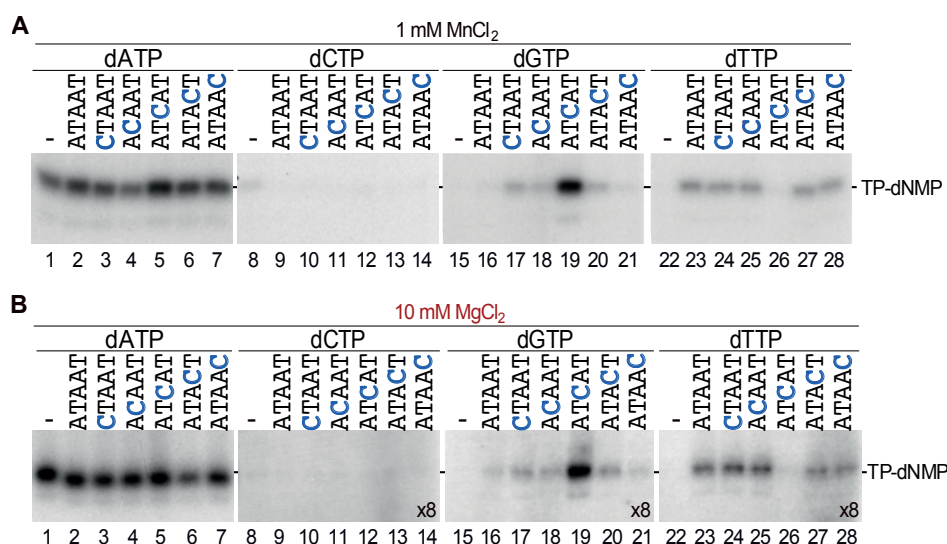
Cuando el TP-DNA fue usado como molde (**Figura 29B**), la incorporación de dAMP fue de nuevo más eficiente que los otros dNMPs (carriles 1-2 y 9-10 frente a los carriles sobreexpuestos 3-8 y 11-16). Sin embargo, con ambos metales la presencia de TP-DNA dio lugar a una disminución del dAMP que se unía a la TP (carriles 1 frente al 2 y 9 frente al 10) y, en el caso de manganeso este efecto también fue observado cuando se utilizó como nucleótido dCTP (carriles 3 frente a 4). Por su lado, la incorporación de dTMP fue detectada con manganeso pero no con magnesio y únicamente en presencia de TP-DNA (carriles 7-8 y 15-16). Estos resultados sugieren que la incorporación de dTMP estaba dirigida por el TP-DNA mientras que los productos obtenidos con dATP y dCTP podrían deberse a una reacción independiente de molde que se podría ver reducida en presencia del TP-DNA debido a una unión de la TP o la B35DNAP al molde de una manera no productiva. La ausencia de incorporación de dTMP con magnesio confirma que, con o sin molde, el manganeso es más eficiente como metal en la iniciación.



**Figura 29. Especificidad de deoxinucleótido en la reacción de iniciación con la TP de Bam35.** A) Iniciación sin molde con los distintos dNTPs. La reacción se activó con MnCl<sub>2</sub> (carriles 1-8) o con MgCl<sub>2</sub> (carriles 9-16) y se incubó durante los tiempos indicados. B) Iniciación con el TP-DNA de Bam35 como molde. La reacción se llevó a cabo con 75 ng de TP-DNA de Bam35 y se incubó 30 min a 37 °C. Los carriles 1-8 muestran el producto de reacción obtenido en presencia de MnCl<sub>2</sub> y los carriles 9-16 con MgCl<sub>2</sub>. En todos los casos el producto TP-dNMP está indicado. Las exposiciones más largas con respecto a la reacción con dATP está indicada para cada deoxinucleótido. Ver Materiales y Métodos 7.1, página 46 para más detalle.

Como se describe en la introducción, la iniciación mediada por TP frecuentemente está dirigida por el segundo, tercer o cuarto nucleótido de la hebra molde del DNA y la replicación completa del genoma necesita de un mecanismo de recuperación de la información genética de las primeras bases (Méndez *et al.*, 1992, Caldentey *et al.*, 1993, King y van der Vliet, 1994). El inicio de la secuencia de la hebra molde de ambos extremos del genoma de Bam35 es 3'-ATAATACCATGGGG..., de esta manera, si la iniciación se produjese frente a la segunda o quinta base, el primer nucleótido incorporado a la TP sería una dAMP y si se produjese frente al primer, tercer o cuarto nucleótido, sería una dTMP. Para analizar estas opciones,

realizamos ensayos de iniciación con un oligonucleótido de banda simple con la secuencia de los primeros 29 nucleótidos de la hebra molde del origen izquierdo del genoma de Bam35 y con cada uno de los cuatro dNTPs. Además, para determinar exactamente la posición del nucleótido que dirige la iniciación, utilizamos ese mismo molde modificado en la primera, segunda, tercera, cuarta o quinta base por una citosina, puesto que la citosina no está presente en las primeras seis posiciones del molde. Como control, se volvió a analizar la deoxinucleotidilación en ausencia de molde. En la **Figura 30A** y **B** (carriles 1-7) se observa que, de nuevo, el dATP fue el deoxinucleótido más eficientemente incorporado con todos los moldes utilizados, tanto con manganeso como con magnesio. Por otro lado, al añadir dCTP (carriles 8-14 en ambos paneles), sólo se detectó la deoxinucleotidilación de la TP poco eficiente en ausencia molde y con manganeso (panel A carril 8), indicando que el dCTP no puede ser usado para la iniciación de la replicación con ninguno de esos sustratos. Por el contrario, cuando se adicionó dGTP (carriles 15-21), al igual que antes, no se observó incorporación en ausencia de molde pero cuando en la tercera posición del oligonucleótido había una citosina se detectó una banda intensa correspondiente al complejo TP-dGMP. La formación de este complejo se puede detectar con ambos metales pero más eficazmente con manganeso (carril 15 frente 19 en ambos paneles). Además, cuando la citosina estaba localizada en la primera, segunda y cuarta posición se obtuvo un producto minoritario de iniciación (carriles 17, 18 y 20). Estos resultados indican que la tercera posición de la hebra molde es la que principalmente determina el deoxinucleótido que se une a la TP en el paso de iniciación. Este resultado se confirmó al analizar el resultado con dTTP (panel A y B carriles 22-28). En este



**Figura 30. Determinación de la posición del molde que dirige la iniciación.** La imagen muestra los resultados de la iniciación con 1 mM de  $\text{MnCl}_2$  (panel A) y con 10 mM de  $\text{MgCl}_2$  (panel B), en ausencia de molde (carriles 1, 8, 15 y 22 de ambos paneles) o en presencia de oligonucleótidos de 29-mer de cadena sencilla con la secuencia del origen izquierdo del genoma de Bam35 (carriles 2, 9, 16 y 23) o con variaciones en la secuencia (demás carriles). En azul se marca el nucleótido modificado con respecto a la secuencia *wild type*. Los deoxinucleótidos utilizados en cada caso y los 6 primeros nucleótidos de cada secuencia (en dirección 3'-5') están indicados encima de los geles. Las exposiciones más largas con respecto a la reacción con dATP está indicadas para cada deoxinucleótido. El protocolo del ensayo y la secuencia completa de los oligonucleótidos están descritos en Materiales y Métodos 7.1, página 46.

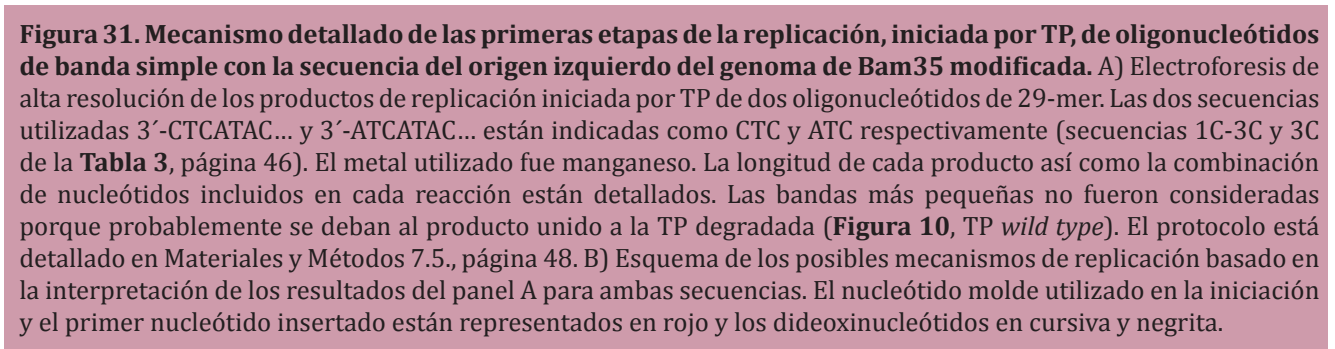
caso, la banda correspondiente al complejo TP-dTMP fue observada con todos los moldes excepto cuando había una citosina en la tercera posición indicando que la presencia de este nucleótido en esa posición impide la iniciación (carriles 23-25 y 27-28 frente a 26 en ambos paneles). Así, tanto por la aparición de producto (cuando se utilizó dGTP) como por la pérdida del mismo (con dTTP) al modificar la tercera base de la secuencia del genoma de Bam35, se puede sugerir que esa base del oligonucleótido es la que dirige la iniciación. Por otro lado, puesto que con TP-DNA la iniciación se produce con dTTP (**Figura 29B** carril 8) y en su tercera posición hay una adenosina, podemos concluir que la iniciación de la replicación del genoma de Bam35 está dirigida por la tercera base de la hebra molde del mismo.



### **5. La información del extremo del genoma de Bam35 puede ser recuperada mediante un nuevo mecanismo de *jumping-back***

Los genomas que se replican utilizando una proteína como *primer* suelen iniciar la replicación en una posición interna del genoma de manera que, para evitar la pérdida de la información genética de sus extremos 3', suelen poseer repeticiones terminales que permiten una translocación hacia atrás del complejo de iniciación TP-(dNMP)<sub>x</sub> (Méndez *et al.*, 1992). Basándonos en la secuencia de los orígenes del genoma de Bam35 (3'-ATAATA...) y en los resultados obtenidos, podemos hipotetizar que la iniciación frente a la tercera posición de la hebra molde debería de ser seguida por un *jumping-back* del complejo formado por la TP unida a un único nucleótido (dTMP) y la B35DNAP hasta la primera base para recuperar la información genética de los extremos del genoma. Para comprobar este mecanismo, analizamos paso por paso la replicación de oligonucleótidos mediante la adición de dideoxinucleótidos en el ensayo (Materiales y Métodos 7.5, página 48). Puesto que, como se observa en la **Figura 30**, la iniciación con dGTP con el oligonucleótido que tiene una citosina en la tercera posición fue más eficaz que con dTTP en la secuencia *wild type* (carriles 19 frente a 23 del panel A y B), decidimos usar un oligonucleótido con la secuencia del origen izquierdo del genoma de Bam35 con la primera y la tercera base modificadas por citosina, 3'-CTCATA... al que llamamos CTC en la **Figura 31** (secuencia 1C-3C de la **Tabla 3**, página 46). Además, utilizamos otro molde en el cual la primera y la tercera base eran distintas con el fin de impedir el mecanismo de recuperación de información genética del extremo, favoreciendo así la copia directa de la secuencia molde, 3'-ATCATA... indicada como ATC en **Figura 31** (secuencia 3C en la **Tabla 3**, página 46). La reacción de iniciación fue más eficiente con el molde ATC que con el CTC y, gracias a la alta resolución de este tipo de gel, detectamos dos bandas con ambos moldes (**Figura 31A** carril 2 frente a 1). Esto podría deberse a una inserción errónea de otro dGMP después de la formación del complejo TP-dGMP. Cuando se añadió un dideoxinucleótido para permitir un paso más en la replicación, se observó un incremento en la intensidad de la banda TP-(dNMP)<sub>2</sub> en comparación con la misma banda observada en el paso de iniciación. Esta banda posiblemente corresponde a los productos TP-dGddA en el caso del molde CTC y TP-dGddT con el molde ATC (carriles 3 y 4). Al añadir dos dNTPs y un









## Capítulo 3.

# ESTUDIO DE LA CAPACIDAD DE REPLICACIÓN A TRAVÉS DE DAÑOS DE LA DNA POLIMERASA DE BAM35



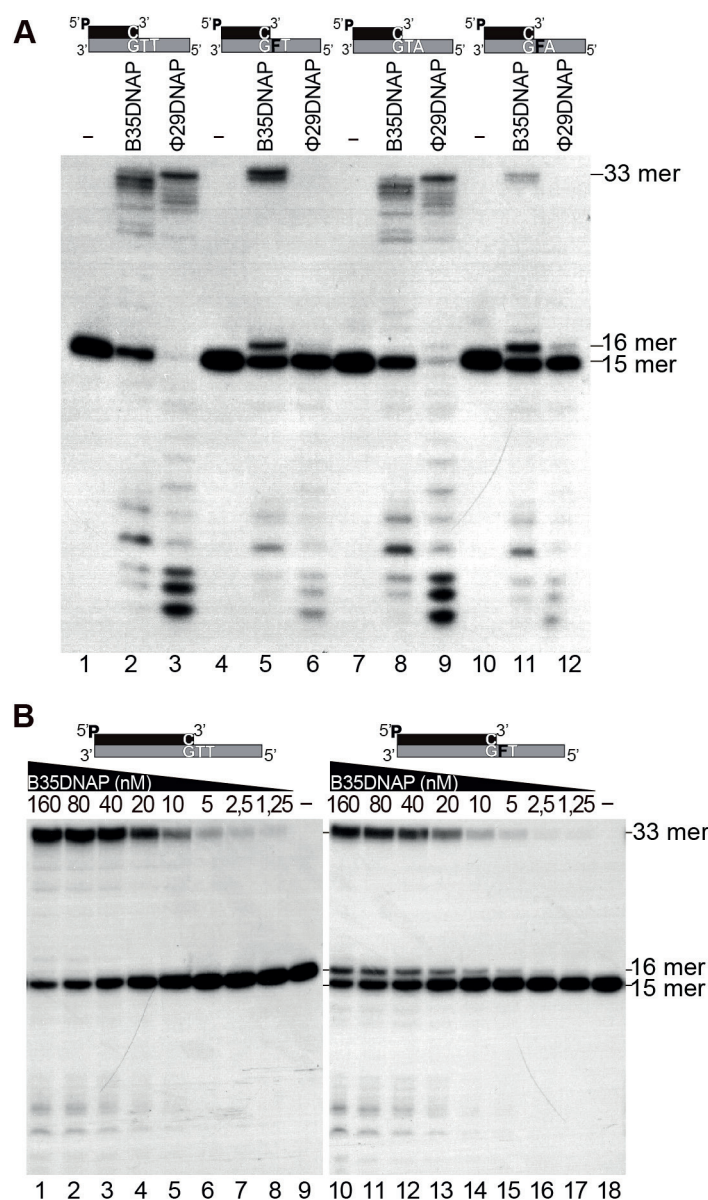
La integridad del DNA está continuamente comprometida por la acción de agentes mutagénicos ambientales capaces de causar diferentes tipos de daños (Liu *et al.*, 2016). Generalmente, la replicación de un DNA dañado requiere de una DNAP con tolerancia a la inserción de un nucleótido que no cumpla las reglas de apareamiento de Watson y Crick, como es el caso de las DNAPs de la familia Y, o de una DNAP con la capacidad de generar errores en el alineamiento entre el *primer* y la hebra molde, lo que produciría mutaciones de cambio de marco de lectura de 1 o 2 nucleótidos como en el caso de la Pol II de la familia B (Wang y Yang, 2009, Goodman y Woodgate, 2013). Sin embargo, el éxito de las DNAPs replicativas para copiar un DNA con daños bloqueantes, por ejemplo sitios abásicos, es muy bajo y depende del contexto de secuencia y de su actividad 3'-5' exonucleasa (Hogg *et al.*, 2004, Tanguy Le Gac *et al.*, 2004, Zhu *et al.*, 2008, Choi *et al.*, 2010). A pesar de esto, el estudio de la capacidad de la B35DNAP replicativa para copiar un molde de DNA con sitios abásicos despertó nuestro interés porque el genoma del bacteriófago Bam35 y otros fagos del grupo podrían estar afectados por agentes mutagénicos endógenos y exógenos como consecuencia de las sucesivas rondas de replicación necesarias para su continuación dentro de la bacteria hospedadora durante su ciclo lisogénico. Además, el ciclo lítico de Bam35 es activado por la respuesta SOS de la bacteria al daño en el DNA (Verheust *et al.*, 2003), existiendo la posibilidad de que el agente causante de esta respuesta también dañase el genoma del bacteriófago. Debido a estas posibilidades, el bacteriófago Bam35 y otros del grupo podrían requerir de una DNAP capaz de replicar un genoma dañado para su supervivencia.



### 1. La B35DNAP es capaz de replicar un molde de DNA con sitios abásicos insertando preferentemente una adenina enfrente del daño

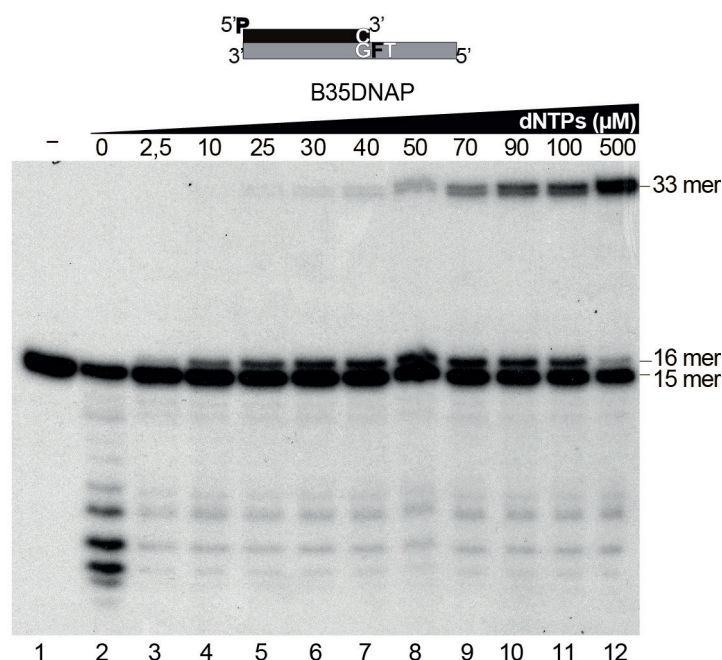
Para estudiar la capacidad de la B35DNAP para sortear sitios abásicos, realizamos un ensayo de replicación de oligonucleótidos con dos moldes que contenían un tetrahidrofurano (THF, o simplemente F), análogo estable de sitio abásico, en dos contextos de secuencia diferente. En ambos moldes el THF se encontraba en la primera posición de replicación, sin embargo, el segundo nucleótido era una T o una A (M33GFT, en sentido 3'-5', **Figura 32A** carriles 4-6 o M33GFA, sentido 3'-5', carriles 10-12); además, se analizaron en paralelo los correspondientes controles sin daño (M33GTT, carriles 1-3 y M33GTA, carriles 7-9). En ausencia de daño, la B35DNAP, al igual que la  $\Phi$ 29DNAP, utilizada como control, fue capaz de extender el *primer* (carriles 3 y 9 frente a 2 y 8). Sin embargo, ante la presencia del THF, la  $\Phi$ 29DNAP fue incapaz de replicar el molde en las condiciones probadas (carriles 6 y 12), mientras que la B35DNAP extendió el *primer* en ambos contextos de secuencia, siendo más efectiva en el contexto de GFT que en GFA (carril 5 frente al 11). Con ambos sustratos la B35DNAP dio lugar a una acumulación de un producto de 16-mer que corresponde con la inserción de un único nucleótido, lo que sugiere que la extensión de este producto es el paso limitante. Además, la replicación por la B35DNAP del molde con daño, M33GFT (**Figura 32B**

carriles 10-18) al igual que el de sin daño, M33GTT (carriles 1-9), fue procesiva, ya que la relación entre el producto de inserción de un nucleótido (producto de 16-mer) y el producto de tamaño completo del molde con THF (33-mer) fue la misma a las diferentes concentraciones de la B35DNAP analizadas. Estos resultados indican que la B35DNAP es capaz de replicar completamente un molde con un sitio abásico sin disociarse del sustrato *primer*/molde.



**Figura 32. Replicación por la B35DNAP de dos moldes de DNA con sitios abásicos.** A) Análisis de los productos de replicación sintetizados por la B35DNAP y la  $\Phi 29$ DNAP al utilizar como molde oligonucleótidos con dos contextos de secuencia (carriles 1-6 frente a 7-12) en ausencia (M33GTT, carriles 1-3 y M33GTA, carriles 7-9) o presencia (M33GFT, carriles 4-6 y M33GFA, carriles 10-12) de un análogo de sitios abásicos (THF, F). Las reacciones se incubaron 30 min a 37 °C. B) Estudio de la procesividad de replicación de la B35DNAP sobre los sustratos C15/M33GTT (carriles 1-9) y C15/M33GFT (carriles 10-18). El ensayo se realizó con un exceso (900 nM) de los sustratos a las concentraciones de B35DNAP indicadas y se incubaron 30 min a 37 °C. Los ensayos del panel A y B se realizaron en presencia de 100 nM de dNTPs para los moldes sin THF y de 100  $\mu$ M para los moldes con THF y los productos se analizaron en gel desnaturante de poliácridamida. En la parte superior de los geles se muestra una representación esquemática de cada sustrato *primer*/molde donde se indica el último nucleótido apareado con el *primer* y los dos primeros nucleótidos molde (GTT, GTA, GFT o GFA), además de las posiciones de los productos de 15, 16 y 33-mer.

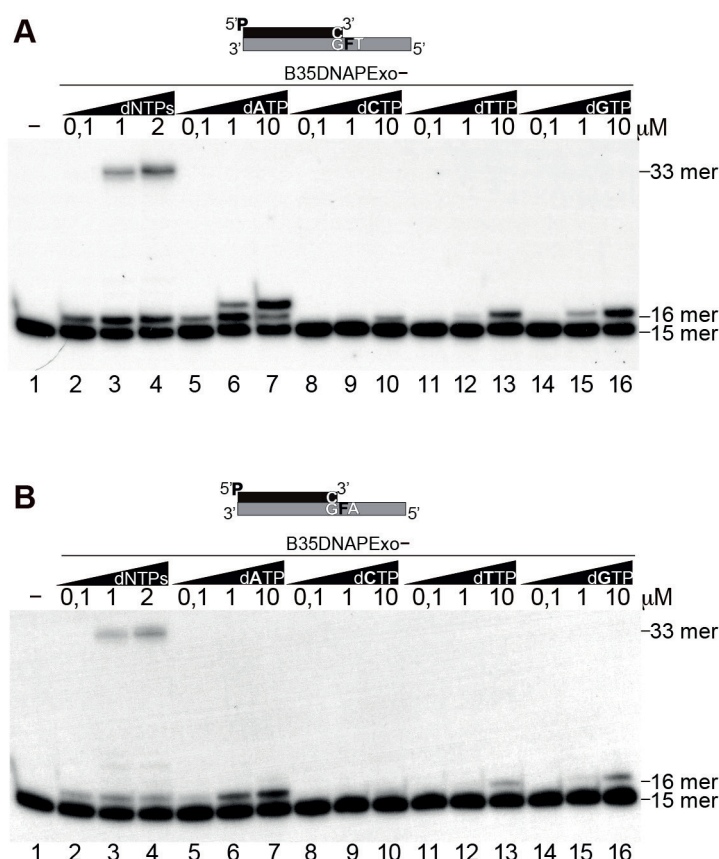
Por otro lado, mediante el estudio del balance de las actividades exonucleasa y polimerasa de la B35DNAP sobre el oligonucleótido molde con THF (M33GFT), se determinó que la concentración mínima de dNTPs para replicar el molde completo era de 40  $\mu\text{M}$  (**Figura 33**, carril 7). Mientras que a concentraciones más bajas (2,5-25  $\mu\text{M}$ ) sólo se observó el producto de 16-mer (carriles 3-5).



**Figura 33. Balance de las actividades exonucleasa y polimerasa de la B35DNAP sobre un molde que presenta un sitio abásico.** El ensayo se realizó con las concentraciones de dNTPs indicadas durante 10 min a 37 °C. Las bandas correspondientes al 15, 16 y 33-mer están indicadas a la derecha. En la parte superior del gel se muestra una representación esquemática del sustrato C15/M33GFT donde se indica el último nucleótido apareado con el primer y los dos primeros nucleótidos molde.

Algunas DNAPs pueden replicar un DNA con sitios abásicos utilizando como molde algún nucleótido posterior al daño, resultando en un producto con uno o dos nucleótidos menos que la longitud total de la cadena molde (Ohashi *et al.*, 2000, Wang y Yang, 2009). En el caso de la B35DNAP el producto de replicación del molde con daño fue de la misma longitud que el obtenido con el molde sin daño, sugiriendo que esta enzima inserta un nucleótido enfrente del THF en lugar de utilizar una base posterior como molde (**Figura 32**). Para confirmar este mecanismo, analizamos la preferencia de incorporación de nucleótido enfrente del daño utilizando la B35DNAPexo- para evitar que la actividad exonucleasa 3'-5' de la B35DNAP eliminase el nucleótido insertado. Así, la **Figura 34** muestra que la B35DNAPexo- inserta preferentemente purinas enfrente del daño, siguiendo el orden de preferencia A>G>T>C, este resultado está en concordancia con la "regla de la A" previamente descrita para la mayoría de las DNAPs (Strauss, 2002). Además, esta preferencia fue la misma para los dos contextos de secuencia analizados (M33GFT, panel A y M33GFA, panel B), lo que descarta la posibilidad de que la B35DNAP esté utilizando el nucleótido posterior al THF como molde.





**Figura 34. Preferencia de inserción de nucleótido enfrente del sitio abásico por la variante B35DNAPexo-.** El ensayo se llevó a cabo con las concentraciones indicadas de cada uno de los dNTPs y con dos contextos de secuencia diferentes, M33GFT (panel A) y M33GFA (panel B) durante 10 min a 37 °C. Los resultados se analizaron en gel de poliacrilamida desnaturalizante. En la parte superior de los geles se muestra una representación esquemática de cada sustrato *primer*/molde utilizado indicando el último nucleótido apareado con el *primer* y los dos primeros nucleótidos molde (GFT o GFA).

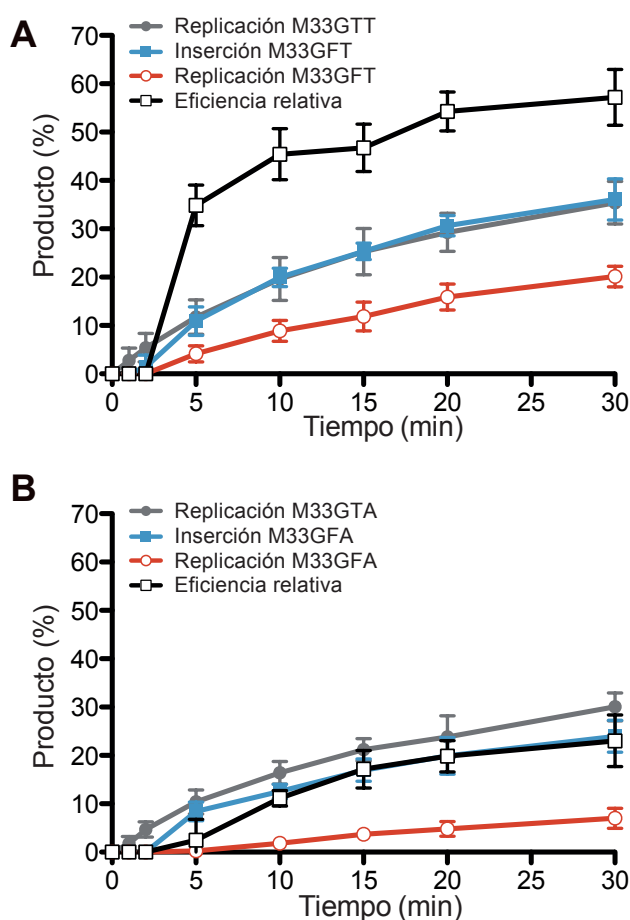


## 2. La secuencia de DNA influye en la eficacia de inserción del nucleótido enfrente del sitio abásico y en la elongación posterior del *primer*

Para profundizar en el estudio de la capacidad de la B35DNAP para replicar un molde con THF, realizamos cinéticas en condiciones de exceso de sustrato (relación 90:1 sustrato:B35DNAP) para favorecer que el producto de polimerización se debiese a la unión de una única molécula de enzima (**Figura 35**). A partir de estos experimentos se calculó, utilizando las ecuaciones indicadas en Materiales y Métodos 8.1 (página 49): la probabilidad de inserción del nucleótido enfrente del THF (línea azul), la probabilidad de la replicación completa del molde con THF (línea roja) y del molde sin THF (línea gris) y la eficiencia relativa de la replicación del sustrato con THF con respecto a la replicación del molde sin THF (línea negra) (Kokoska *et al.*, 2003). La **Figura 35A** muestra el resultado obtenido con los moldes M33GTT y M33GFT. En este caso, la probabilidad de inserción del nucleótido enfrente del THF (línea azul) siguió la misma cinética que la



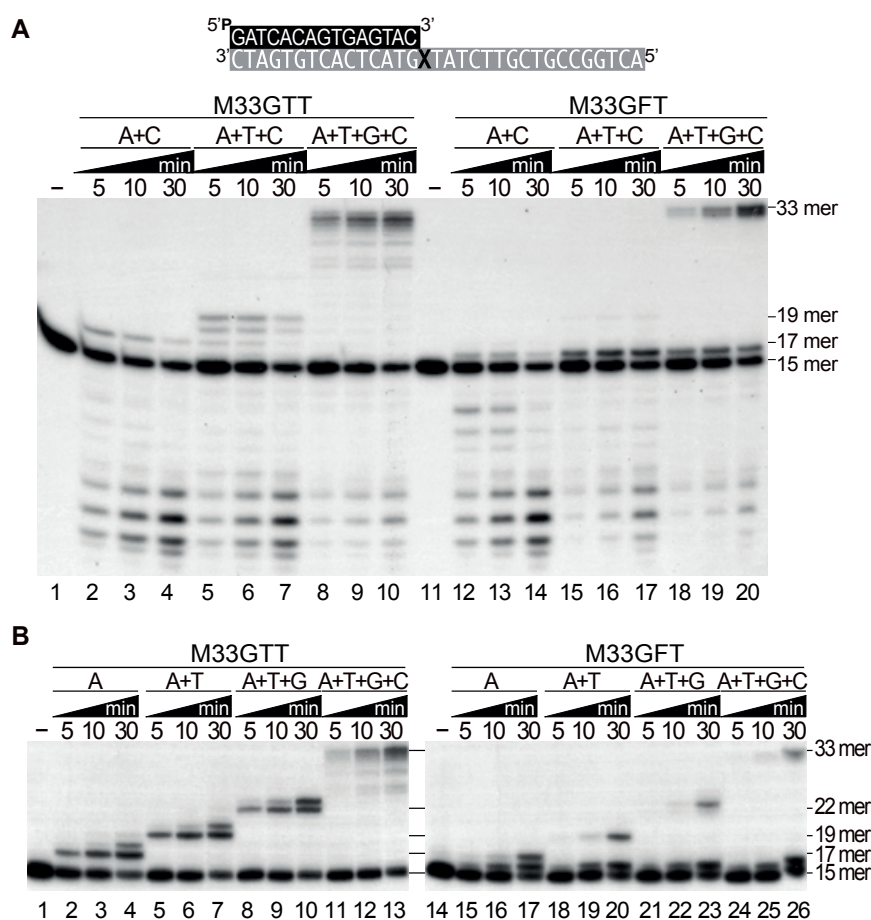
probabilidad de la replicación completa del molde sin THF (línea gris). Sin embargo, la probabilidad de la replicación completa del molde con THF (línea roja) fue mucho menor, lo que sugiere que la inserción del primer nucleótido es muy eficaz siendo su posterior extensión el paso limitante. La eficiencia relativa de replicación entre M33GFT/M33GTT (línea negra) fue de aproximadamente un 60%. Por el contrario, con el molde M33GFA (Figura 35B), la probabilidad de inserción del nucleótido enfrente del THF (línea azul) fue menor que la probabilidad de la replicación completa del molde sin THF (línea gris), lo cual, junto al paso limitante de elongación (línea roja), resultó en una eficiencia relativa de replicación inferior al 20% (línea negra). Este resultado indica que el contexto de secuencia influye, tanto en la capacidad de inserción del nucleótido enfrente del THF como en la extensión posterior del *primer*.



**Figura 35. Análisis a lo largo del tiempo de la replicación de dos moldes de DNA con sitios abásicos por la B35DNAP.** Los ensayos se realizaron con un relación 90:1 de sustrato:B35DNAP, con los oligonucleótidos molde M33GTT y M33GFT (gráfica A) o M33GTA y M33GFA (gráfica B) y 100  $\mu$ M de dNTPs. Se muestra la media y DE de tres experimentos independientes. La línea azul representa la probabilidad de inserción de un nucleótido enfrente del THF; la línea roja la probabilidad de la replicación completa del molde con THF; la gris la probabilidad de replicación del molde sin THF y la negra la eficiencia relativa de replicación entre el molde con THF y el molde sin daño. Las ecuaciones utilizadas se muestran en Materiales y Métodos 8.1, página 49

### 3. La actividad correctora de pruebas de la B35DNAP contrarresta la replicación de oligonucleótidos con sitios abásicos hasta que el daño se encuentra a más de 5 nucleótidos de distancia del extremo 3' del primer

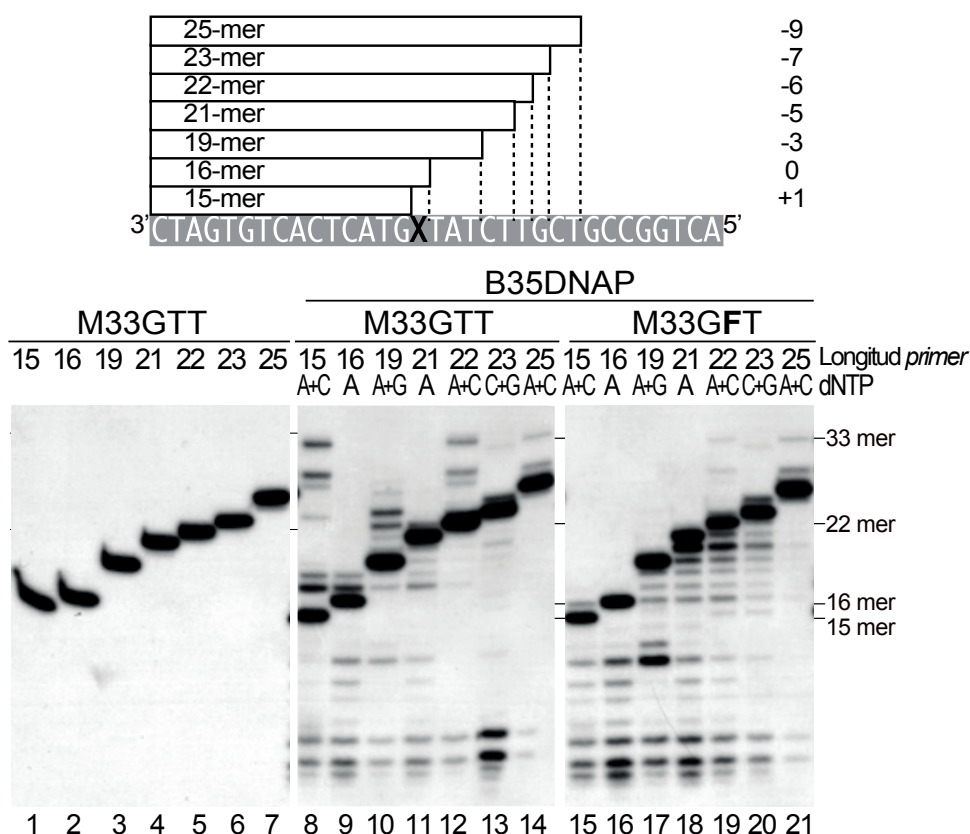
Con el objetivo de profundizar más en el estudio de la replicación de moldes de DNA con sitios abásicos por la B35DNAP, analizamos la extensión del primer paso a paso. Para ello, realizamos un ensayo de replicación de oligonucleótidos en presencia de dCTP, para minimizar la degradación del nucleótido del extremo 3' del primer, y de dATP, dATP+dTTP o dATP+dTTP+dGTP (Materiales y Métodos 8.1, página 49). En la **Figura 36A** se observa que, en ausencia de daño (M33GTT), al ir añadiendo las diferentes combinaciones de dNTPs las bandas obtenidas correspondían con la elongación del primer siguiendo un correcto apareamiento, observando un producto de 17-mer en presencia de A proveniente de la replicación del molde 3'-TT (dos nucleótidos más que el primer, carriles 2-4), de 19-mer con A+T correspondiente al molde 3'-TTAT (cuatro nucleótidos más, carriles 5-7) y la replicación completa en presencia de los 4 dNTPs (33-mer, carriles 8-10). Sin embargo, con el molde M33GFT, la adición de A dio lugar a un producto de 16-mer en lugar



**Figura 36. Estudio secuencial de la replicación de un oligonucleótido con sitios abásicos por la B35DNAP y su variante sin actividad exonucleasa.** A) Polimerización por la B35DNAP de un sustrato primer/molde sin (carriles 1-10) o con THF (carriles 11-20). Con el molde sin THF se utilizó una concentración de 100 nM de dNTPs y con THF de 100 µM. Las combinaciones de dNTPs en cada caso están indicadas. En la parte superior del gel se representa un esquema del sustrato primer/molde utilizado; el primer fue el C15, y la X del molde indica la posición de una timidina o del THF (M33GTT o M33GFT). B) Polimerización de la B35DNAPexo- de un molde sin o con THF (carriles 1-13 o 14-26, respectivamente). La concentración de las diferentes combinaciones de dNTPs utilizadas fue de 50 nM para el molde sin THF y 1 µM con THF. A la derecha de cada gel se indica el tamaño de los distintos productos obtenidos.

de 17-mer. Esta banda corresponde a la inserción de un único nucleótido enfrente del THF, pero el par A:THF, en lugar de elongarse (puesto que el siguiente nucleótido es una T), era degradado por la actividad exonucleasa (carriles 12-14). La elongación tampoco se observó en presencia de A+T (carriles 15-17) siendo únicamente posible en presencia de los 4 dNTPs (carriles 18-20). Estos resultados sugieren que cuando la A es insertada enfrente del THF, la DNAP detecta la falta de apareamiento de manera similar a un error en el apareamiento, lo que da lugar a ciclos fútiles de degradación y polimerización que sólo puede ser superado cuando los 4 dNTPs están disponibles, lo que le permite insertar más nucleótidos de manera procesiva después de replicar enfrente del THF. Esto fue confirmado mediante la realización del mismo experimento con la B35DNAPexo- (**Figura 36B**). En este caso, el molde con THF dio lugar a los productos esperados tras 30 min de reacción, aunque la acumulación del producto de 16-mer es igualmente detectado, remarcando la dificultad de elongar el *primer* más allá del sitio abásico. Este resultado está en concordancia con lo observado en las cinéticas permitiéndonos concluir que el paso limitante en la replicación de oligonucleótidos con THF es la extensión del nucleótido no apareado enfrente del THF, ya que este paso se ve fuertemente contrarrestado por la actividad exonucleasa 3'-5'.

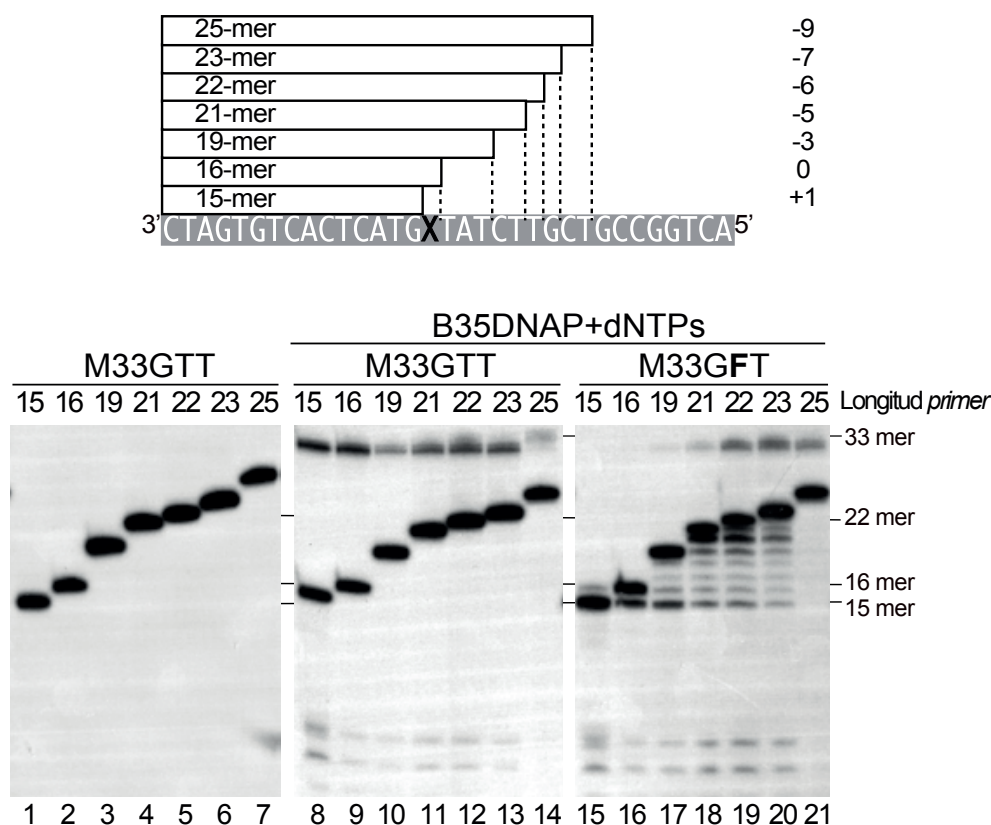
Para analizar este proceso con más detalle, realizamos ensayos de replicación de oligonucleótidos con sustratos *primer*/molde en los que utilizamos distintos *primers* para que su extremo 3' estuviese hibridado en distintas posiciones posteriores al THF (Materiales y Métodos 8.1, página 49), en presencia de una alta concentración (100  $\mu$ M) del primer dNTP complementario al molde y del dNTP correspondiente al extremo 3' del *primer* para disminuir su degradación. En ausencia de daño el nucleótido correcto fue insertado en todos los sustratos, aunque también se observaron algunas bandas correspondientes a productos de replicación erróneos debido a la alta concentración de dNTPs (**Figura 37**, carriles 8-14). Sin embargo, con el molde que contiene un THF en la posición 16 (M33GFT), solamente se detectaron productos de degradación del *primer* cuando este tenía una longitud de 15-21 nucleótidos, lo que corresponde con la posición del THF en +1, 0, -3 y -5 con respecto al extremo 3' del *primer* (carriles 15-18). Esto sugiere que el no apareamiento generado por el nucleótido enfrente del sitio abásico es detectado por la B35DNAP como un error en el apareamiento, lo que produciría la transferencia del *primer* a su sitio activo de exonucleasa. Por el contrario, cuando el extremo 3' del *primer* estuvo a 6 o más posiciones del THF (*primer* de 22-mer o superior, carriles 19-21), se observaron bandas de replicación. Esto sugiere que la B35DNAP es capaz de detectar la combinación nucleótido:THF, hasta que se encuentra a 5 posiciones del extremo 3' del *primer* lo que dificultaría su extensión. Sin embargo, a partir de esta distancia, el error no sería detectado permitiendo la replicación del molde.



**Figura 37. Detección de la falta de apareamiento nucleótido:THF por la B35DNAP.** En la parte superior se representa un esquema de los sustratos *primer*/molde utilizados. La X del molde indica la posición de una timidina o THF. La distancia relativa de la X con respecto al extremo 3' del *primer* está anotada a la derecha. Las reacciones se llevaron a cabo durante 10 min a 37 °C en presencia de 10 nM B35DNAP y 100 µM de los dNTPs indicados (último dNTP del *primer*, para evitar su degradación, y siguiente dNTP complementario al nucleótido del molde). En la parte superior del gel se indica la longitud del *primer* y a su derecha las posiciones de las bandas correspondientes a 15, 16, 22 y 33-mer.

Para entender como la B35DNAP detecta el no apareamiento nucleótido:THF, realizamos el mismo ensayo pero con los cuatro dNTPs a una concentración en la que la B35DNAP puede introducir el nucleótido enfrente del THF pero no elongar el *primer* (10 µM, **Figura 33** carril 4). En estas condiciones, observamos que el sustrato *primer*/molde sin daño fue replicado completamente sin importar el tamaño del *primer* (**Figura 38**, carriles 8-14). En el caso del molde con THF, se observó la inserción de un único nucleótido cuando el THF estaba en la posición +1 con respecto al extremo 3' del *primer* (carril 15) y cuando este estuvo enfrente del THF no se observó elongación (carril 16). Incluso cuando el THF estuvo a 3 nucleótidos de distancia del extremo (carril 17) sólo se observó una banda de producto muy tenue. Únicamente cuando el THF estuvo en la posición -5 (carril 18) o más claramente en la -6 o más lejos (carriles 19-21), la replicación del tamaño completo del molde fue detectada, indicando que la B35DNAP es capaz de detectar la combinación nucleótido:THF como un error de apareamiento hasta que el THF está a 4-5 posiciones del extremo 3' del *primer*, en concordancia con los resultados anteriores. Además, se observaron productos de degradación que llegaban hasta el tamaño de 15-mer con todos los *primers* utilizados excepto con el más largo (25-mer)

(carriles 15-20 frente a carril 21). Estos productos pueden corresponder a la degradación exonucleolítica del *primer* hasta el nucleótido enfrente del THF, reafirmando que cuando el error en el apareamiento está a 9 nucleótidos del extremo 3' del *primer* la B35DNAP no es capaz de detectar el par nucleótido:THF. En conjunto, estos resultados confirman que la B35DNAP es capaz de detectar el daño hasta una distancia de 5 nucleótidos del extremo 3'. Por lo tanto, para que la replicación de oligonucleótidos con sitios abásicos sea exitosa precisa de una síntesis procesiva de DNA.



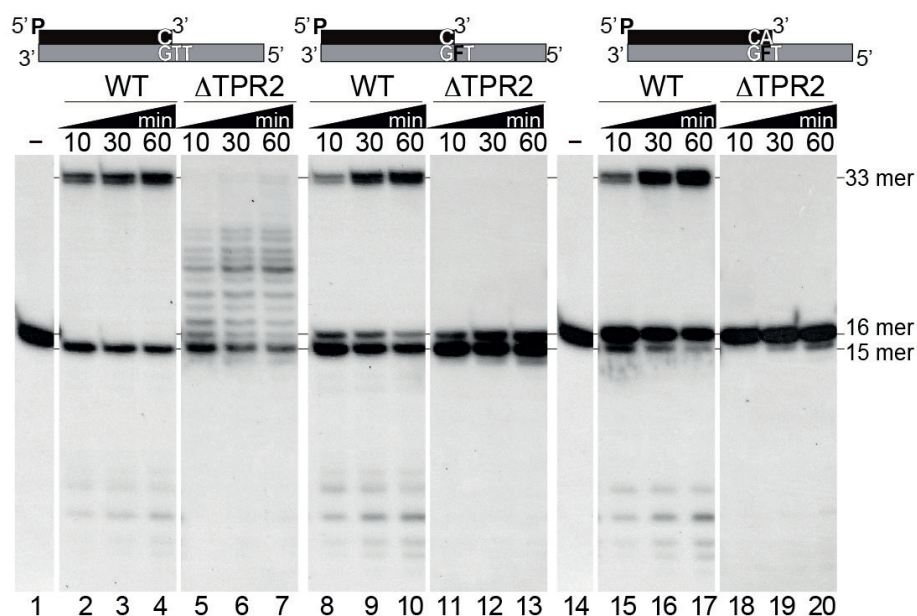
**Figura 38. Detección del no apareamiento nucleótido:THF por la B35DNAP.** En la parte superior se representa un esquema de los sustratos *primer*/molde utilizados. La X del molde indica la posición de una timidina o THF. La distancia relativa de la X con respecto al extremo 3' del *primer* está anotada a la derecha. Las reacciones se llevaron a cabo durante 10 min a 37 °C en presencia de 10 nM B35DNAP y 10 µM de dNTPs. En la parte superior del gel se indica la longitud del *primer* y a su derecha las posiciones de las bandas correspondientes a 15, 16, 22 y 33-mer.

#### 4. La eliminación del subdominio TPR2 de la B35DNAP no afecta a la inserción del nucleótido enfrente del THF, pero impide la extensión posterior del *primer*

Estudios previos sugieren que los subdominios de las DNAPs necesarios para la procesividad podrían estar implicados en la replicación de DNAs con sitios abásicos. Así, por ejemplo, en un estudio se determinó que el subdominio TPR2 de una DNAP de la familia B derivada de un polintón que inicia la replicación con una TP era requerido para la replicación de moldes con daños (Pastor-Palacios *et al.*, 2012). Por otro lado, la capacidad de replicación a través de sitios abásicos por la Pol θ humana dependió de un único motivo que



también es requerido para la replicación y procesividad (Hogg *et al.*, 2011). Puesto que el subdominio TPR2 de la B35DNAP es necesario para la procesividad (**Figura 15**), quisimos determinar su papel en la replicación de moldes con sitios abásicos. Para ello, comparamos la replicación, llevada a cabo por la variante  $\Delta$ TPR2 de la B35DNAP, de un oligonucleótido sin THF (M33GTT) con el DNA con THF (M33GFT). Así, observamos que la proteína  $\Delta$ TPR2 era capaz de replicar el molde sin daño (**Figura 39**, carriles 5-7) pero que ante la presencia del THF en la primera posición del molde (carriles 11-13), sólo era capaz de insertar un nucleótido enfrente del daño, con una eficacia similar a la proteína *wild type* (alrededor de un 25% en 10 min). Sin embargo, no se detectó la elongación del sustrato más allá del THF ni siquiera a los tiempos más largos de incubación utilizados. Para confirmar este resultado, quisimos analizar si, al proporcionar un *primer* con un nucleótido enfrente del THF (C16, 16-mer), la proteína  $\Delta$ TPR2 era capaz de elongarlo y dar lugar a una replicación completa del molde. En estas condiciones la B35DNAP *wild type* replicó el oligonucleótido en sólo 10 min (**Figura 39**, carril 15), mientras que la  $\Delta$ TPR2 fue incapaz de extender el *primer* y sólo se detectó, a una hora de reacción, una banda de degradación de 15-mer (carriles 18-20). Estos resultados indican que la incapacidad mostrada por la proteína  $\Delta$ TPR2 para replicar oligonucleótidos con THF podría deberse a la pérdida de procesividad, que no obstante no parece esencial para introducir un nucleótido enfrente del daño. En conjunto, podemos concluir que la procesividad de la B35DNAP contribuye



**Figura 39. Estudio de la replicación de un oligonucleótido con sitios abásicos por la variante  $\Delta$ TPR2 de la B35DNAP.** Actividad de polimerización de la B35DNAP *wild type* (10 nM) y su variante  $\Delta$ TPR2 (200 nM) con oligonucleótidos de DNA sin o con THF. Las reacciones se incubaron a 37 °C durante los tiempos indicados en presencia de 100  $\mu$ M de dNTPs. En la parte superior de los geles se muestra un esquema de cada sustrato *primer*/molde utilizado indicando los últimos nucleótidos del *primer* y los dos primeros nucleótidos molde (GFT). Las posiciones de los oligonucleótidos de 15, 16 y 33-mer están indicados a la derecha.

a su capacidad para replicar DNA con sitios abásicos al facilitar el alcance de una distancia de 5 nucleótidos con respecto al daño, distancia a la cual la “memoria” de la DNAP ya no puede detectarlo permitiendo, por tanto, la elongación del *primer* a través del daño.







## Capítulo 4.

# MAPEO DE LAS INTERACCIONES ENTRE LAS PROTEÍNAS CODIFICADAS POR EL GENOMA DE BAM35



Tal y como se comenta en la Introducción, la determinación de la función de las proteínas codificadas por el genoma de los fagos es muy complicada debido a su gran diversidad genética. En el caso particular de Bam35, únicamente se ha demostrado o predicho la función de 17 proteínas de las 32 codificadas por su DNA (**Figura 4**). Esta información no pudo ser ampliada mediante la comparación de las secuencias aminoacídicas de las proteínas de Bam35 con nuevas bases de datos como *Phagonaute* o *pVOGs* debido a la ausencia de similitudes significativas (Delattre *et al.*, 2016, Grazziotin *et al.*, 2016). Sin embargo, puesto que el conocimiento de la función de las proteínas tiene una gran importancia en el entendimiento de la biología de los fagos, uno de los objetivos de este capítulo fue aportar nuevos indicios que pudiesen ayudar a la identificación de las funciones desconocidas de las proteínas codificadas por el genoma de Bam35.

Por otro lado, como se ha mencionado anteriormente, se ha descrito que la presencia de proteínas adicionales *in vitro* incrementa la eficiencia de replicación del DNA de algunas DNAPs siendo estas proteínas esenciales para la replicación *in vivo*, como por ejemplo las proteínas de unión a DNA de banda doble (P6) y de banda simple (P5) del bacteriófago  $\Phi 29$  (Blanco *et al.*, 1986, Martín *et al.*, 1989) o, las proteínas de unión a banda simple P12 y P19 de PRD1 (Pakula *et al.*, 1990, Pakula *et al.*, 1993). Aún más, en el caso de adenovirus o del bacteriófago Nf (Dekker *et al.*, 1997, Longás *et al.*, 2006), estas proteínas son esenciales para la replicación *in vitro*. Por tanto, el segundo objetivo de este capítulo fue la búsqueda de otras proteínas que interaccionasen con la B35DNAP o con la TP, como las P1 y P2, putativas proteínas de unión a DNA, y que eventualmente pudiesen incrementar la eficiencia de la replicación *in vitro* del TP-DNA de Bam35 (**Figura 24**).

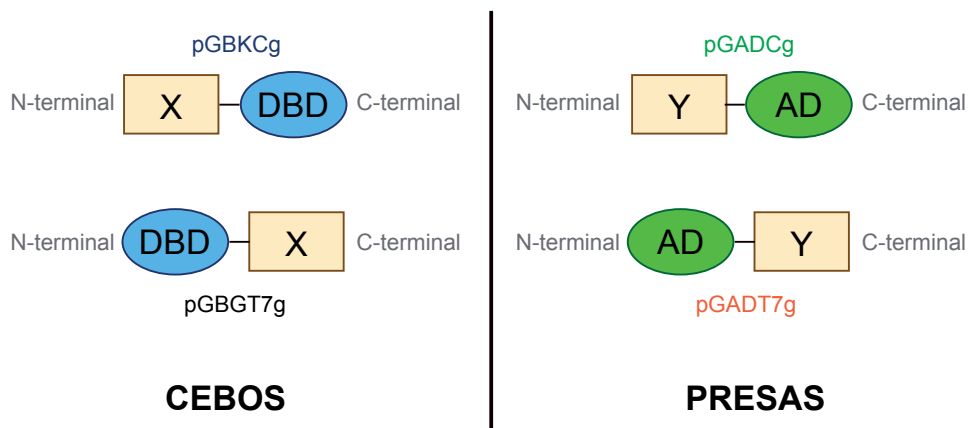
Para cumplir ambos objetivos, decidimos realizar un análisis de la interacción entre todas las proteínas codificadas por el genoma de Bam35 utilizando el sistema de doble híbrido en levaduras, que ha sido de mucha utilidad para el estudio de otros fagos, especialmente en cuanto a proteínas estructurales (Rajagopala *et al.*, 2011, Mehla *et al.*, 2015b). El estudio mostrado en este capítulo se llevó a cabo en colaboración con el Dr. Uetz (Virginia Commonwealth University, Richmond).

En este capítulo, con el fin de simplificar la nomenclatura, utilizamos solo el nombre de la proteína viral para referirnos a las proteínas de fusión codificadas por los vectores indicados.

## 1. Obtención de librerías de expresión de todos los ORFs de Bam35 en cuatro vectores de levaduras

Debido al hecho de que al utilizar diferentes combinaciones de plásmidos para realizar el sistema de doble híbrido se obtienen distintos conjuntos de interacciones (Rajagopala *et al.*, 2011, Osterman *et al.*, 2015), decidimos clonar los 32 ORFs del genoma de Bam35 en cuatro vectores diferentes (**Figura 40** y Materiales y Métodos 9, página 52). De esta manera, el plásmido cebo pGBGT7g fusiona el dominio DBD del factor de

transcripción Gal4 al extremo N-terminal (N) de la proteína clonada, mientras que el pGBKCg lo hace al extremo C-terminal (C). Por su lado, el plásmido presa pGADT7g fusiona el dominio AD de Gal4 al extremo N-terminal (N) de la proteína de interés y el pGADCg lo hace al extremo C-terminal (C). Al combinar cada uno de los cebos con ambas presas nos permite analizar la interacción entre las proteínas de fusión en todas las conformaciones posibles (NN, NC, CN y CC) evitando perder información debido a restricciones estructurales.

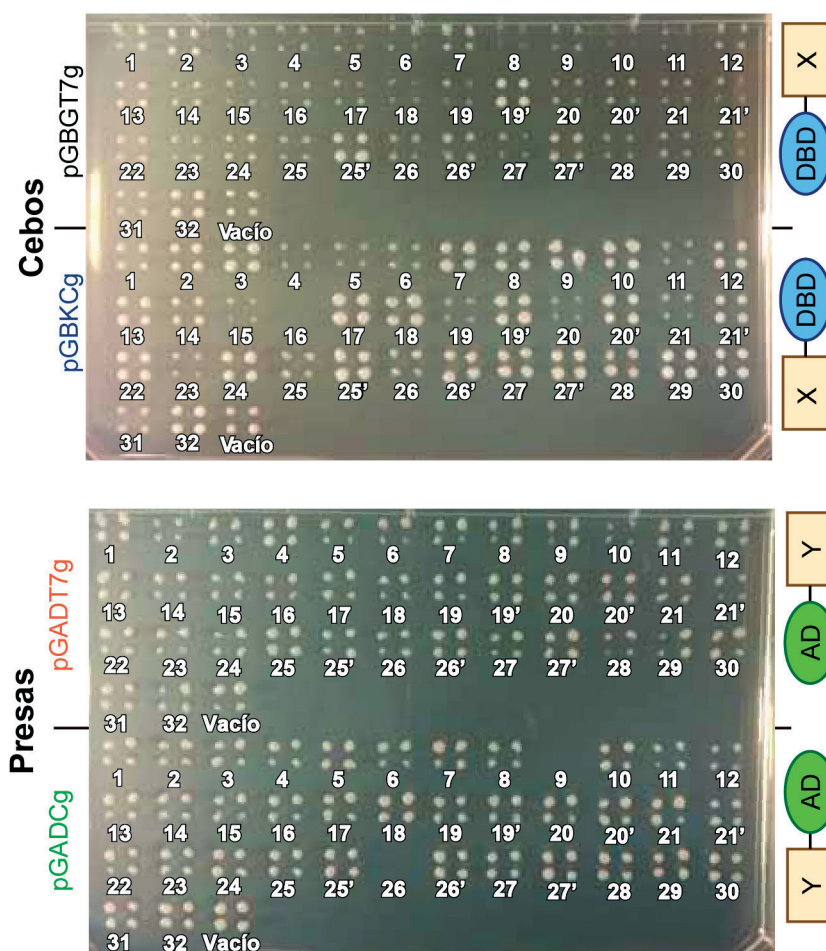


**Figura 40. Esquema de las proteínas de fusión codificadas por los vectores de expresión en levaduras utilizados en esta tesis.** Las proteínas cebo (X) están fusionadas al dominio DBD del factor de transcripción Gal4, en posición C-terminal en el vector pGBKCg (C), o N-terminal en pGBGT7g (N). Las proteínas presa (Y) están fusionan al dominio AD de Gal4, en posición C-terminal en pGADCg (C), o N-terminal en pGADT7g (N).

Por otro lado, utilizamos el programa TMHMM (Krogh *et al.*, 2001) para reevaluar las predicciones de posibles motivos transmembrana de las proteínas codificadas por los ORFs del genoma de Bam35. De los 32 ORFs, 9 (ORF15, 16, 19, 20, 21, 23, 25, 26 y 27) codificaban proteínas con alta probabilidad de poseer dichos dominios los cuales podrían afectar a la solubilidad de la proteína expresada y, por tanto, evitar la interacción con otras proteínas. Por ello, decidimos generar variantes de esos 9 ORFs eliminando la secuencia que codificaba el dominio de transmembrana predicho, creando así una versión corta de cada uno de ellos (indicada con un '). En los ORFs 15, 16 y 23, el dominio de transmembrana ocupaba gran parte de la proteína, por lo que únicamente diseñamos la construcción con su versión completa.

La construcción del ORF16 en pGBKCg (C) no fue posible y, al seleccionar las levaduras transformadas con las construcciones que contenían los ORFs 9 y 26 en pGADCg (C) no se obtuvo crecimiento (**Figura 41**). Por lo tanto, finalmente, obtuvimos todas las construcciones en los vectores de expresión de levaduras pGADT7g (N) y pGBGT7g (N), todas en pGADCg (C) excepto las que contenían los ORFs 9 y 26 y todas en pGBKCg (C) a excepción del ORF16. Al comprobar la secuencia nucleotídica que codificaba cada proteína, encontramos que 5 proteínas, en 3 clones diferentes, tuvieron mutaciones puntuales que modificaron el aminoácido codificado: la proteína P8 tuvo su aminoácido 20 modificado de arginina a treonina; la proteína P10 cambió la fenilalanina 89 por una isoleucina; la P14 la alanina 172 a prolina; la P17 la serina 55 a

arginina y la P18 la asparagina 69 a isoleucina. Puesto que se detectó exactamente la misma mutación en los 3 clones secuenciados y además se había utilizado una DNAP de muy alta fidelidad para la amplificación de los ORFs por PCR (Kapa HiFi), no podemos descartar que la secuencia de la base de datos sea errónea. Debido a esto y a que un único aminoácido probablemente no afecte a gran parte de las interacciones detectables, decidimos continuar, cautelarmente, el análisis con esas proteínas.



**Figura 41.** Levaduras transformadas con los vectores cebos pGBGT7g o pGBKCg, o los vectores presas pGADT7g o pGADCg con los distintos ORFs de Bam35 clonados. Las construcciones en los plásmidos cebo se utilizaron para transformar la cepa haploide AH109 y se seleccionaron en medio sin triptófano. Las presas se emplearon para transformar la cepa haploide Y187 y se seleccionaron en medio sin leucina. Se obtuvieron todas las construcciones excepto los ORFs 9 y 26 en pGADCg y el ORF16 en pGBKCg. Los números indican el ORF del genoma de Bam35 presente en cada vector. Las versiones cortas se indican con '.



## 2. Determinación de las condiciones óptimas de selección

Para evitar falsos positivos debido a la inducción del crecimiento de las levaduras por la biosíntesis de histidina provocada por la activación del promotor UAS únicamente por la proteína de fusión cebo (auto-activación) en lugar de por la interacción entre las proteínas cebo-presa, analizamos el resultado obtenido del apareamiento de cada uno de los clones de la cepa haploide AH109 transformada con cada uno de los vectores cebo (pGBKCg (C) o pGADT7g (N)) con los ORFs de Bam35 con los clones de la cepa

haploide Y187 transformada con el plásmido vacío pGADT7g o pGADCg. Después de seleccionar los clones diploides mediante el plaqueo en medio sólido sin leucina ni triptófano, se analizó el crecimiento en placas de medio sin leucina, triptófano ni histidina a las que se adicionaron diferentes concentraciones de un inhibidor competitivo de la histidina, 3AT (ver Materiales y Métodos, 9.3, página 56), con el objetivo de incrementar la cantidad de producto del gen reportero HIS3 necesario para el crecimiento de las levaduras. Así, establecimos la concentración mínima necesaria de 3AT para evitar el crecimiento por auto-activación para cada una de las proteínas cebo (**Tabla 8**). En todos los casos, al aparear la levadura transformada con el mismo ORF clonado en uno de los vectores cebos (ya sea en N- o en C-terminal) con la levadura portadora del plásmido vacío pGADT7g necesitaron la misma concentración de 3AT que con la portadora del plásmido vacío pGADCg. Además, las proteínas cebo expresadas por el plásmido pGBKCg (C) mostraron un mayor grado de auto-activación que las codificadas por pGBGT7g (N), necesitando una mayor concentración de 3AT para evitar el crecimiento de las colonias. Así, con pGBKCg tan sólo 10 clonajes no mostraron crecimiento por auto-activación mientras que 20 necesitaron una concentración superior a 25 mM de 3AT para evitar el fondo. Sin embargo, con pGBGT7g 22 clonajes no dieron lugar a auto-activación y tan sólo 3, de los cuales 2 correspondieron a versiones cortas de las proteínas, necesitaron una concentración de 3AT superior a 25 mM.

| Concentración 3AT      | pGBGT7g (N) + pGADT7g o pGADCg vacío  | pGBKCg (C) + pGADT7g o pGADCg vacío           |
|------------------------|---|---|
| 0 mM                   | 1, 3, 4, 5, 6, 8, 9, 11, 15, 16, 18, 19, 20, 20', 21', 23, 25, 26, 27, 28, 29, 31 | 4, 5, 6, 11, 15, 20, 21, 23, 25, 31           |
| 25x10 <sup>-3</sup> mM | 13, 14, 17, 21, 22, 24, 26'   | -   |
| 0,1 mM                 | 30  | 26  |
| 3 mM                   | 2, 7, 10, 12, 27', pGBGT7g vacío  | 1, 14, 28                                     |
| 10 mM                  | -   | 7, 8, 26', pGBKCg vacío                       |
| 25 mM                  | 19', 25'  | 10, 19, 20', 29                               |
| 50 mM                  | 32  | 2, 3, 9, 12, 13, 17, 18, 21', 22, 24, 27, 27' |
| 100 mM                 | -   | 19', 25', 30, 32                              |

**Tabla 8. Concentraciones de 3AT mínimas necesarias para evitar la auto-activación de las proteínas cebo.** Se apareó cada clon de la cepa AH109 (plásmidos cebo, pGBGT7g (N) y pGBKCg (C) con los ORFs de Bam35) con la cepa Y187 transformada con el vector vacío pGADT7g o pGADCg (indicados en la parte superior). Los clones diploides se crecieron en medio sólido sin leucina, triptófano ni histidina y con distintas concentraciones de 3AT (primera columna). Los ORFs están incluidos en las celdas correspondientes a la concentración mínima de 3AT necesaria para evitar su auto-activación. Las versiones cortas de los ORFs se indican con '.



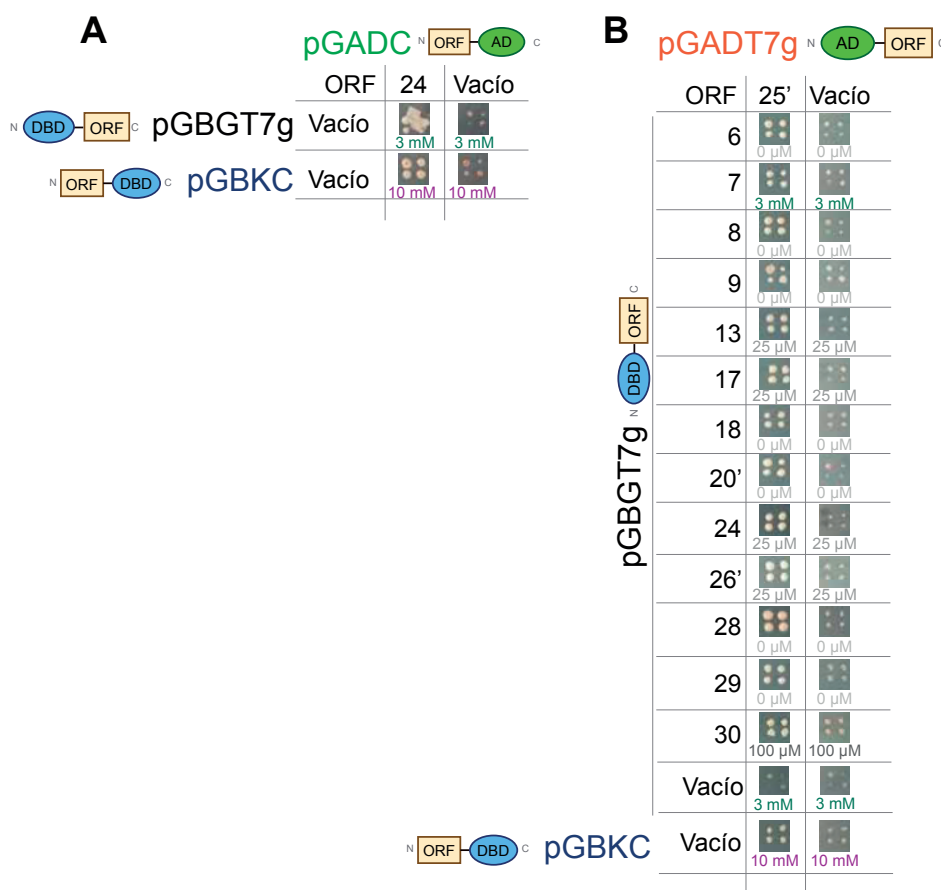
### 3. Interactoma proteico de Bam35

#### 3.1. Establecimiento de los criterios seguidos para la interpretación de los resultados.

##### Eliminación de los posibles falsos positivos

Para determinar todas las interacciones proteína-proteína del bacteriófago Bam35 se aparearon cada

uno de los clones de las levaduras transformadas con los plásmidos cebo, pGBGT7g (N) y pGBKCg (C) que contienen cada uno de los ORFs del genoma de Bam35, con cada uno de los clones transformados con los plásmido presa, pGADT7g (N) y pGADCg (C) que contienen también cada uno de los ORFs. Después de seleccionar los clones diploides en medio sin leucina ni triptófano, analizamos las interacciones mediante el crecimiento de las levaduras en medio sin leucina, triptófano ni histidina. Inicialmente se realizó un análisis utilizando la concentración de 3AT de referencia establecida en el apartado anterior (**Tabla 8**). En este primer cribado, se observó que las levaduras diploides que contenía la construcción pGADCg::ORF24 (C) y el vector vacío pGBGT7g o el vector vacío pGBKCg fueron capaces de crecer en presencia de la concentración de 3AT establecida para evitar el crecimiento por auto-activación en las levaduras transformadas con las combinaciones de vectores vacío pGADCg y pGBGT7g o pGADCg y pGBKCg, es decir, 3 mM y 10 mM respectivamente (**Figura 42A**), lo que indicaba que la proteína presa P24 estaba activando la expresión del gen HIS3 por si sola dando lugar a un falso positivo. Por otro lado, la versión sin dominio de transmembrana



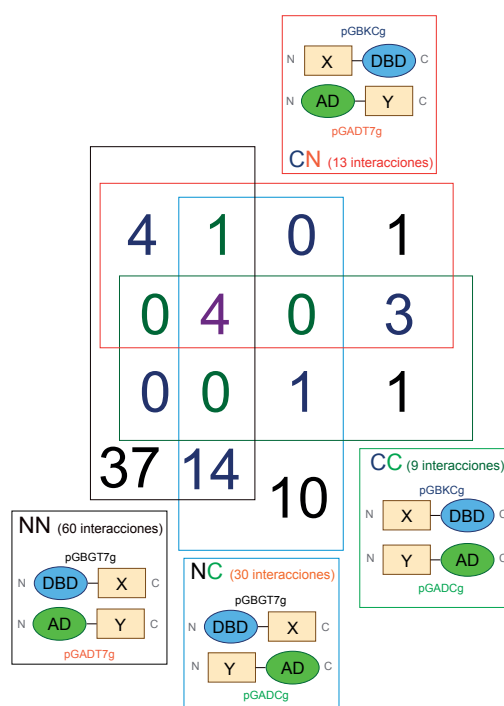
**Figura 42. Detección de la auto-activación de la proteína presa P24 y posible falso positivo de la proteína P25'.**

A) Fotografía representativa del crecimiento de las levaduras transformadas con los vectores vacíos pGBGT7g o pGBKCg y la construcción pGADCg::ORF24 a 3 y 10 mM, respectivamente. Como control se muestra el crecimiento de las levaduras transformadas con los vectores vacíos pGBGT7g o pGBKCg y el vector vacío pGADCg a las mismas concentraciones de 3AT. B) Fotografías representativas de las interacciones encontradas entre la proteína presa P25' expresada a partir del vector pGADT7g (N) y las proteínas cebo de Bam35 expresadas por pGBGT7g (N) o el plásmido vacío pGBKCg, a las concentraciones de 3AT indicadas. Como control se muestra el crecimiento de las levaduras que expresan las mismas proteínas cebo a partir del plásmido pGBGT7g (N) o el pGBKCg vacío y el vector vacío pGADT7g a las mismas concentraciones de 3AT.



de la proteína P25 (P25') codificada por el vector pGADT7g (N) fue muy promiscua en la interacción con las proteínas cebo expresadas a partir del plásmido pGBGT7g (N), porque, aunque no se observó crecimiento por auto-activación de la proteína presa P25' a 3 mM 3AT, dio lugar a 13 interacciones. Sin embargo, en la levadura diploide que portaba los vectores pGADT7g::ORF25' y el vector vacío pGBKCG si se detectó crecimiento debido a la auto-activación de la proteína presa P25' a 10 mM de 3AT (**Figura 42B**). Debido a que esta proteína es una versión modificada de la proteína P25 de Bam35, a su alta promiscuidad y a la auto-activación observada decidimos catalogar estas interacciones como falsos positivos.

Una vez descartados los posibles falsos positivos, con el primer cribado se obtuvieron un total de 129 interacciones, las cuales fueron sometidas a un nuevo análisis pero, en este caso, a diferentes concentraciones de 3AT (0, 25, 100  $\mu$ M y 3, 10, 25 y 50 mM) para poder determinar la fuerza de cada una de las interacciones. Al mismo tiempo se analizó la auto-activación de las proteínas cebo (pGBGT7g::ORF (N) y pGBKCG::ORF (C) contra pGADT7g y pGADCg vacíos) y de las presas (pGBGT7g y pGBKCG vacíos frente a pGADT7g::ORF (N) y pGADCg::ORF (C)). Se consideraron interacciones positivas cuando el crecimiento de la levadura expresando dos proteínas de fusión de Bam35 fue mayor que los controles de auto-activación o cuando el crecimiento fue el mismo que los controles, pero las levaduras con ambas proteínas mostraron un crecimiento normal y los controles de auto-activación dieron lugar a colonias con crecimiento estresado (colonias rojizas). Así, un total de 112 interacciones pudieron ser confirmadas, de las cuales 22 (19,5%) fueron observadas con



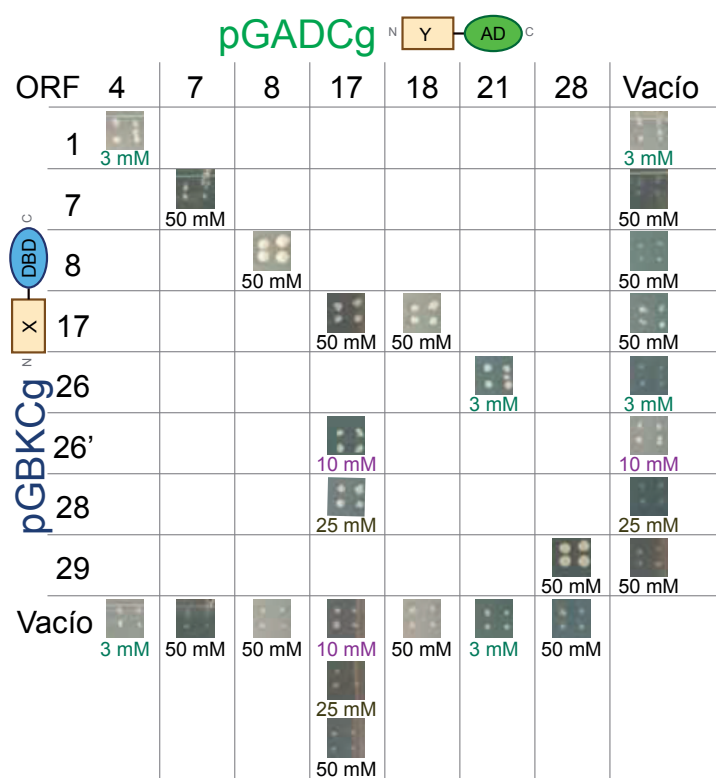
**Figura 43. Número de interacciones encontradas entre las distintas combinaciones (CN, CC, NC y NN) de las proteínas de Bam35 fusionadas a los dominios del factor de transcripción Gal4.** El número en morado corresponde a las interacciones observadas con todas la combinaciones ensayadas. En verde las encontradas en 3 combinaciones diferentes, en azul en 2 y en negro en sólo una combinación. Dentro de los cuadros de color se representa las fusiones utilizadas en cada caso y el número total de interacciones. En total se obtuvieron 112 interacciones, de las cuales 63 se encontraron en más de una combinación.



dos de las versiones cortas de las proteínas (P20' y P26'). Las otras 90 (80,5%) se obtuvieron con un total de 23 proteínas (71,8% del proteoma de Bam35). De las 112 interacciones, 63 se detectaron con más de una combinación de plásmidos, proporcionando 27 interacciones distintas a las que hay que sumar las 49 que sólo se detectaron con una combinación, dejando un total de 76 interacciones diferentes (**Figura 43**). Cuatro de las 76 interacciones (3,5%) se observaron con todas las combinaciones de plásmidos probadas incrementando su fiabilidad (**Figura 43**). De estas 4 interacciones, 3 fueron interacciones de la proteína de fusión consigo misma (proteínas P8, P7 y P17), sugiriendo que esas proteínas podrían formar multímeros, y la otra interacción se obtuvo entre la versión corta de la proteína P26 (P26') y la proteína P17. Una única interacción se detectó con la combinación de 3 vectores distintos, coincidiendo, en este caso, con una proteína de la espícula que interaccionaba consigo misma (proteína P28). Por otro lado, 22 (19,6%) de las interacciones se observaron con la combinación de dos plásmidos y de las 49 (43,7%) obtenidas con una sola combinación, 37 (33%) se detectaron con las proteínas de fusión expresadas por la pareja pGBGT7g/pGADT7g (NN) y tan sólo 1 por la combinación pGBKCg/pGADT7g (CN) y otra con pGBKCg/pGADCg (CC) (**Figura 43**). Estas diferencias en el número de interacciones obtenidas con las distintas combinaciones están en concordancia con los resultados obtenidos con el bacteriófago  $\lambda$  (Rajagopala *et al.*, 2011).

### 3.2. Interacciones detectadas con la combinación de los vectores pGBKCg y pGADCg (CC)

Al analizar las interacciones encontradas al combinar las proteínas de fusión expresadas a partir de los vectores pGBKCg y pGADCg (CC) (**Figura 44**) se observó que todas ellas resistieron altas concentraciones de

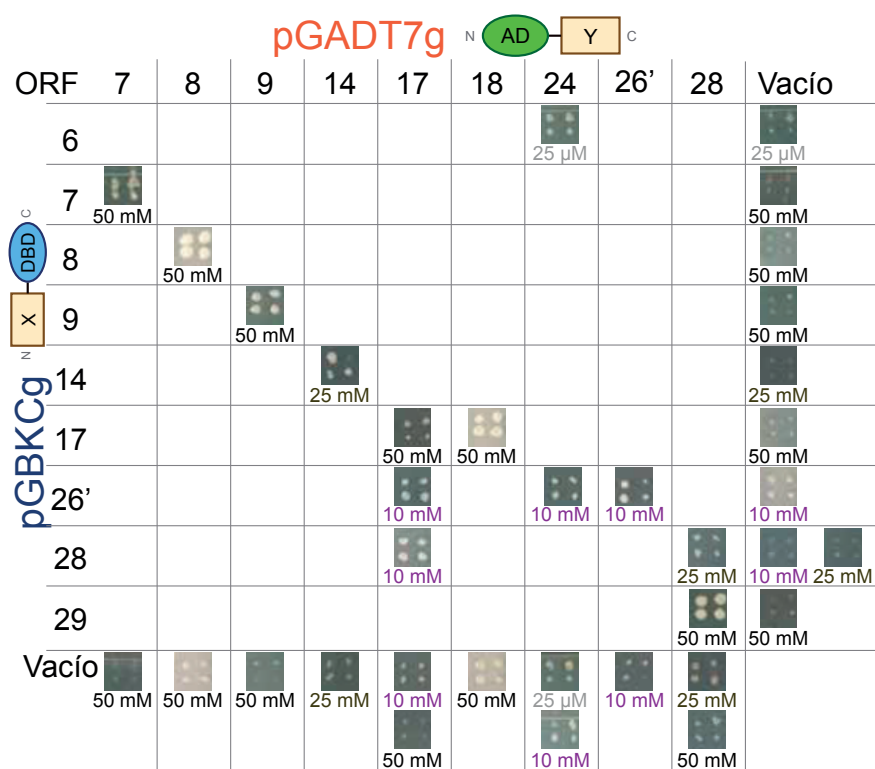


**Figura 44. Interacciones positivas entre las proteínas de Bam35 fusionadas por su extremo C-terminal a los dominios del factor de transcripción Gal4 (CC).** El esquema muestra una fotografía representativa del crecimiento de las levaduras con interacciones positivas entre las proteínas de fusión que expresan a la mayor concentración de 3AT tolerada. Como controles de crecimiento por auto-activación de la proteína cebo se muestra el crecimiento de las levaduras transformadas con el vector pGBKCg::ORF (C) y el vector vacío pGADCg y para la proteína presa con el vector vacío pGBKCg y pGADCg::ORF (C).

3AT. Además, de las 9 interacciones obtenidas con estas proteínas, 8 se encontraron en otras combinaciones de plásmidos, lo que aumenta la fiabilidad del resultado obtenido. Por otro lado, con la combinación CC fue con la única con la que se observó una interacción entre proteínas en las que se ha comprobado o predicho su implicación en la replicación del DNA (a excepción de la interacción de la TP consigo misma). Así, la proteína de unión a DNA (P1) junto a la TP de Bam35 (P4) dieron lugar al crecimiento de las levaduras a una concentración de 3 mM de 3AT.

### 3.3. Interacciones detectadas con la combinación de los vectores pGBKCg y pGADT7g (CN)

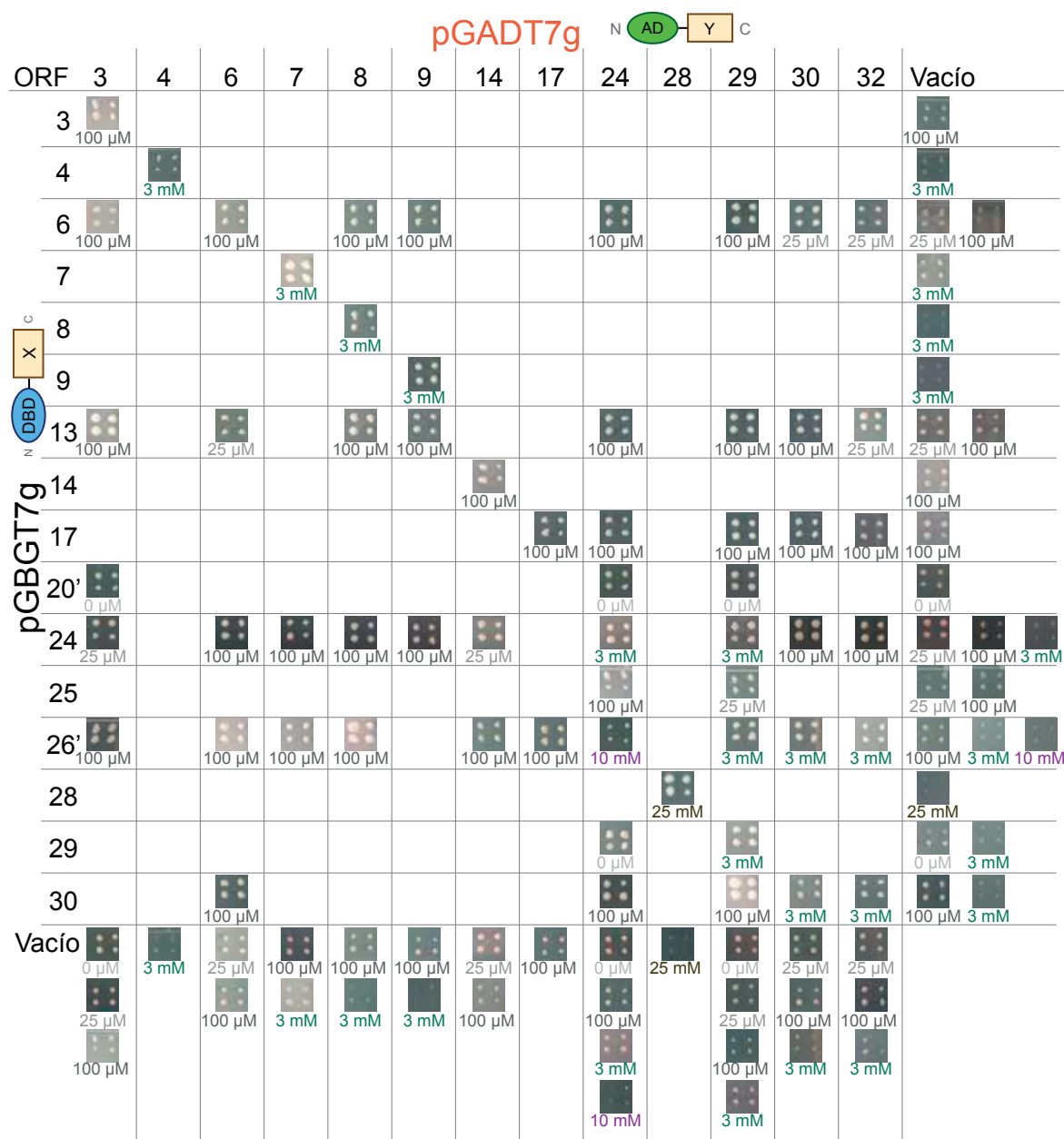
De manera similar, con la combinación entre las proteínas de fusión expresadas a partir de los vectores pGBKCg y pGADT7g (CN) (**Figura 45**) las interacciones observadas toleraron altas concentraciones de 3AT y 12 de las 13 se detectaron con otro tipo de conformación. Cabe destacar que 7 de las 13 interacciones se observaron cuando ambos vectores expresaban la misma proteína de fusión, lo que sugiere que estas proteínas pueden formar multímeros.



**Figura 45. Interacciones positivas entre las proteínas cebo de Bam35 fusionadas en su extremo C-terminal al dominio DBD del factor de transcripción Gal4 y las proteínas presa fusionadas por su N-terminal al dominio AD (CN).** El esquema muestra una fotografía representativa del crecimiento de las levaduras con interacciones positivas entre las proteínas de fusión que expresan a la mayor concentración de 3AT tolerada. Como controles de crecimiento por auto-activación de la proteína cebo se muestra el crecimiento de las levaduras transformadas con el vector pGBKCg::ORF (C) y el vector vacío pGADT7g y para la proteína presa con el vector vacío pGBKCg y pGADT7g::ORF (N).

### 3.4. Interacciones detectadas con la combinación de los vectores pGBGT7g y pGADT7g (NN)

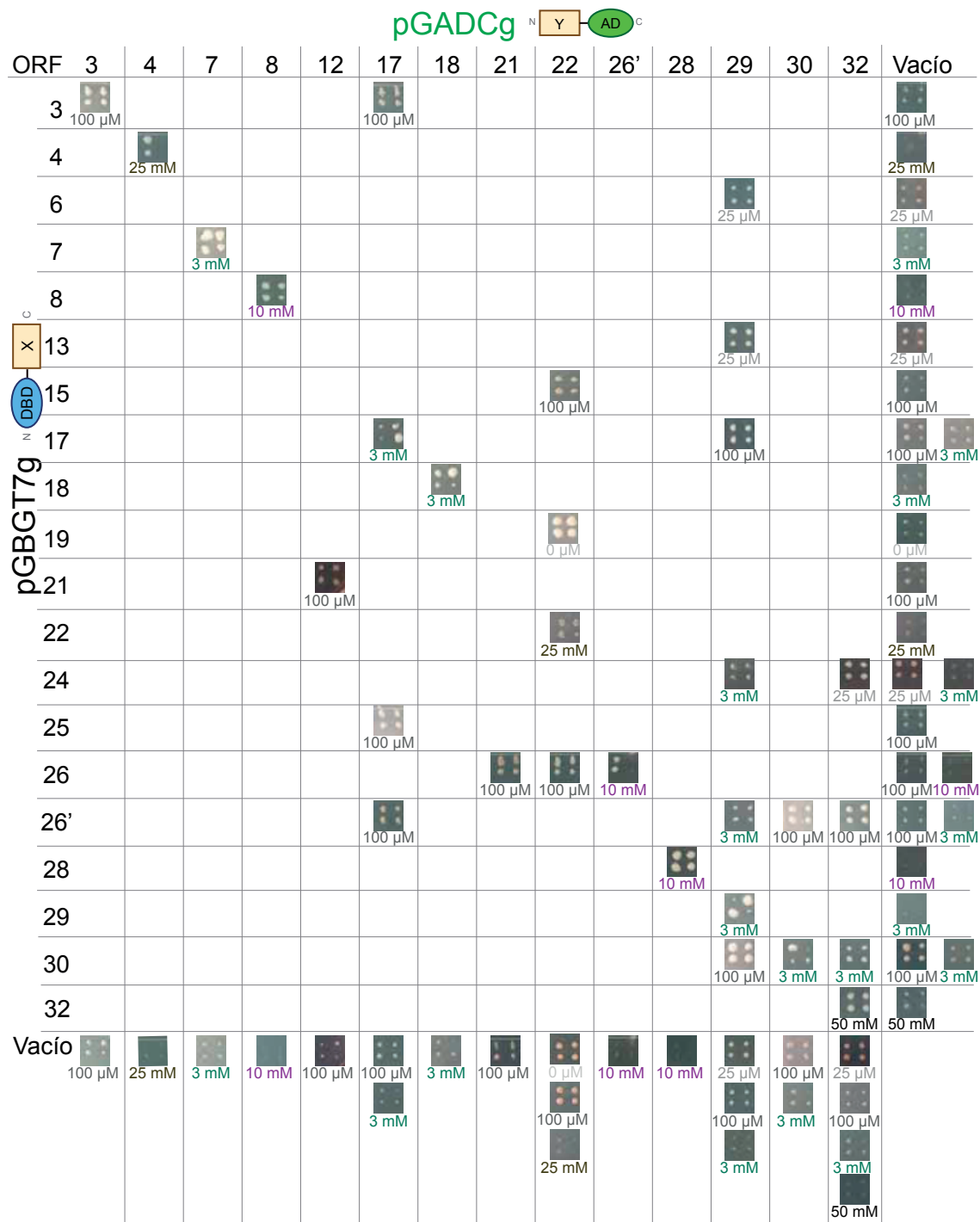
Por otro lado, la combinación de las proteínas clonadas en los vectores pGBGT7g y pGADT7g (NN) (Figura 46), mostraron una tolerancia menor a 3AT. Además, las proteínas cebo P6, P13, P24 y P26' y las proteínas presa P24 y P29 interaccionaron con un gran número de proteínas produciendo 36 y 18 interacciones, respectivamente, frente a las 60 totales. La interacción entre la proteína P26' y P24 fue la más permisiva al 3AT.



**Figura 46. Interacciones positivas entre las proteínas de Bam35 fusionadas por su extremo N-terminal a los dominios del factor de transcripción Gal4 (NN).** El esquema muestra una fotografía representativa del crecimiento de las levaduras con interacciones positivas entre las proteínas expresadas a la mayor concentración de 3AT tolerada. Como controles de crecimiento por auto-activación de la proteína cebo se muestra el crecimiento de las levaduras transformadas con el vector pGBGT7g::ORF (N) y el vector vacío pGADT7g y para la proteína presa con el vector vacío pGBGT7g y pGADT7g::ORF (N).

### 3.5. Interacciones detectadas con la combinación de los vectores pGBGT7g y pGADCg (NC)

A diferencia del resultado obtenido con la pareja de vectores pGBGT7g y pGADT7g, con la combinación entre las proteínas clonadas en pGBGT7g y pGADCg (NC) (**Figura 47**) no se observó ninguna proteína cebo con más de 4 interacciones y sólo la proteína presa P29 dio lugar a 7 interacciones (frente a 30 totales).

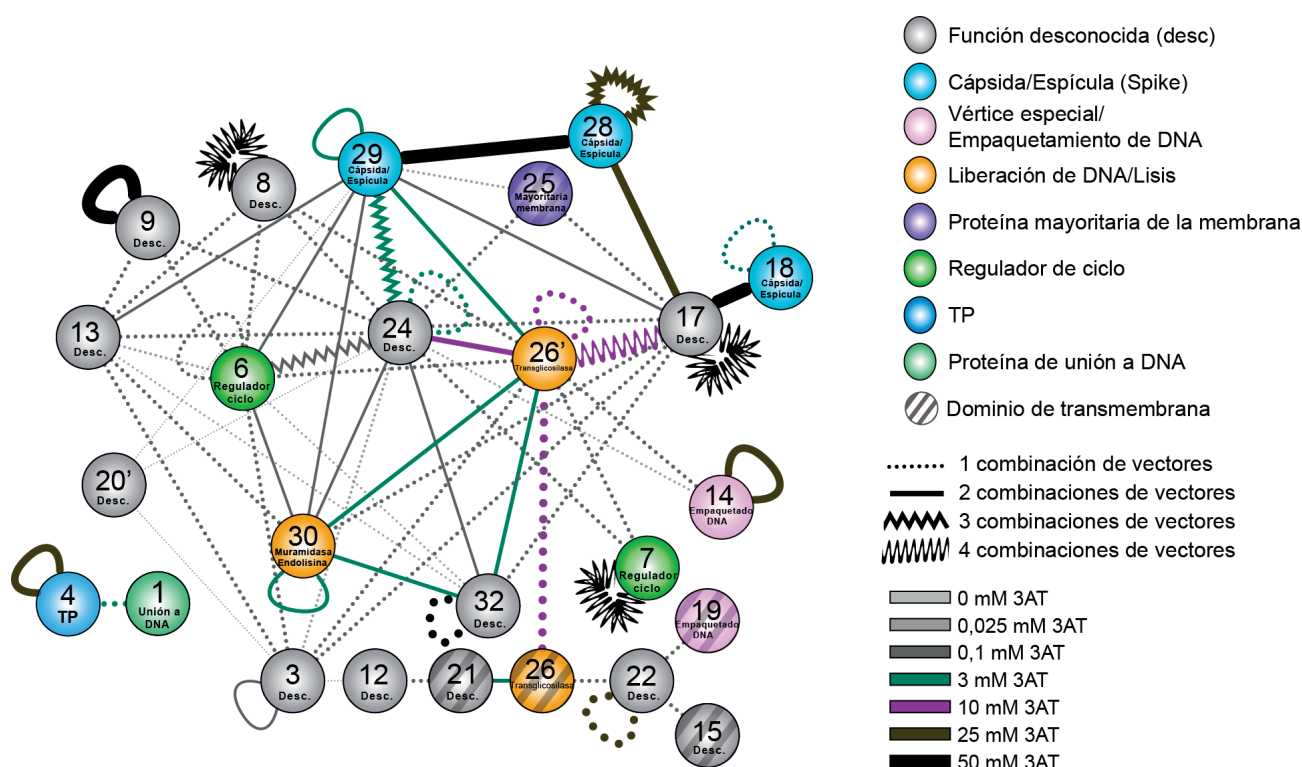


**Figura 47. Interacciones positivas entre las proteínas cebo de Bam35 fusionadas por su extremo N-terminal al dominio DBD del factor de transcripción Gal4 y las proteínas presa fusionadas en su C-terminal al dominio AD (NC).** El esquema muestra una fotografía representativa del crecimiento de las levaduras con interacciones positivas entre las proteínas de fusión que expresan a la mayor concentración de 3AT tolerada. Como controles de crecimiento por auto-activación de la proteína cebo se muestra el crecimiento de las levaduras transformadas con el vector pGBGT7g::ORF (N) y el vector vacío pGADCg y para la proteína presa con el vector vacío pGBGT7g y pGADCg::ORF (C).



#### 4. Interactoma proteico de Bam35

Finalmente, para integrar todos los resultados obtenidos, hicimos un mapa de interacción entre las proteínas codificadas por el genoma de Bam35 en el cual indicamos la función conocida o predicha de las proteínas codificadas por los distintos ORFs. Para resaltar la fuerza de la interacción, representamos las aristas de unión con diferente estilo y color según el número de combinaciones de plásmidos en los que se detectaron y la concentración máxima de 3AT tolerada (**Figura 48**). De las 25 proteínas que interaccionaron, 12 son de función desconocida. Entre las proteínas de función conocida o sugerida, la mayor parte pertenecen a la cápsida o a las espículas de Bam35 o están implicadas en la liberación del DNA o en la lisis de la bacteria. Además, dos de ellas (P6 y P7) se encargan de la regulación del ciclo vital del bacteriófago.



Curiosamente, las dos proteínas de versión corta, P20' y P26' no mostraron la misma interacción que sus versiones completas. La proteína P20 no mostró ninguna interacción, mientras que la P20' interaccionó con las proteínas P3, P24 y P29. Estas diferencias podrían deberse bien a defectos de expresión, de agregación o de estabilidad de la proteína P20 o al hecho de que al quitar el dominio de transmembrana (proteína P20') se haya modificado la estructura terciaria de la proteína lo que podría producir que dominios internos de la proteína ahora se encuentren en la parte externa dando lugar a interacciones no posibles *in vivo*. En el

caso de la proteína P26 presentó interacciones con la proteína P21 y P22, sin embargo, su versión corta mostró interacciones fuertes y probablemente reales, puesto que se detectaron con varias combinaciones de vectores a 3 y 10 mM de 3AT, consigo misma (tanto entre la versión corta-versión corta como versión corta-versión completa) y con las proteínas P17, P24, P29, P30 y P32 además de interacciones más débiles con P3, P6, P7, P8 y P14. Todas las interacciones encontradas se discuten en detalle en el apartado correspondiente de la Discusión.



## DISCUSIÓN

5







## 1. La B35DNAP es una enzima procesiva y fiel

A pesar de pertenecer a familias de bacteriófagos diferentes y tener una morfología viral distinta, las DNAP de  $\Phi 29$  y Bam35 pertenecen a la familia B y tienen una estructura muy similar presentando los mismos dominios y subdominios. Las principales diferencias se encuentran en el subdominio *thumb* y TPR2 que son más largos en el caso de la B35DNAP. Entre las funciones del *thumb* se encuentran la de conferir estabilidad en la unión al extremo del *primer* y coordinar las actividades exonucleasa y polimerasa (Beese *et al.*, 1993, Minnick *et al.*, 1996, Kamtekar *et al.*, 2004). Por otro lado, el motivo TPR2 es esencial para la procesividad y el desplazamiento de banda (Kamtekar *et al.*, 2004, Rodríguez *et al.*, 2009). Gracias a la presencia de esos subdominios, la B35DNAP replica DNA de una manera procesiva acoplada al desplazamiento de banda, características imprescindibles para llevar a cabo la replicación completa del TP-DNA viral sin la necesidad de proteínas adicionales.

En esta tesis se ha comprobado que la B35DNAP tiene una clara preferencia por moldes con más de 10 nucleótidos en el extremo 5' protuberante en ensayos de replicación con oligonucleótidos. Sin embargo, el tamaño del dúplex de DNA formado por el *primer*/molde no afecta a su capacidad de replicación. Estos resultados sugieren que la B35DNAP podría requerir un sustrato de DNA que presente un molde largo para establecer una unión estable. Esta hipótesis está apoyada por el resultado obtenido en el ensayo de retraso en gel, en el cual sólo se observa una unión estable de la B35DNAP cuando el oligonucleótido molde tiene entre 12 y 14 nucleótidos protuberantes en el extremo 5'. Este resultado difiere de la  $\Phi 29$ DNAP, puesto que esta enzima es capaz de replicar moldes pequeños los cuales son retrasados en el gel (Blasco *et al.*, 1993). Nuestros resultados sugieren que la dificultad para replicar un oligonucleótido con pocos nucleótidos protuberantes en el extremo 5' mostrado por la B35DNAP podía deberse, en parte, a su subdominio TPR2 ya que la variante de delección de este subdominio fue capaz de replicar este tipo de moldes. Puesto que el subdominio TPR2 interacciona con la cadena molde (Berman *et al.*, 2007), es probable que la presencia de un TPR2 de mayor tamaño requiera también de un mayor número de contactos con esta cadena para se forme una interacción estable con el DNA.

La presencia de una actividad 3'-5' exonucleasa junto a la alta preferencia de inserción del nucleótido complementario al nucleótido molde mostradas por la B35DNAP hacen que esta enzima tenga una alta fidelidad durante la replicación. Así, la tasa de error de la B35DNAP es de  $5,1 \pm 1,7 \times 10^{-6}$ , rango similar al mostrado por la  $\Phi 29$ DNAP ( $10^{-6}$ ) y otras DNAPs comerciales como la *Pfu* ( $10^{-6}$ ), insertando todas ellos menos errores que la *Taq* polimerasa ( $10^{-5}$ ) (Cline *et al.*, 1996, Nelson *et al.*, 2002, McInerney *et al.*, 2014). En línea con estos resultados, la conservación de las secuencias de los genomas mostrada por los datos disponibles y la estabilidad en la secuencia de una región altamente variable presente

en virus altamente relacionados con Bam35 tras 50 días de rondas de replicación del fago, indican que la replicación de genomas virales por las DNAPs de estos fagos también se produce fielmente *in vivo* (Jalasvuori *et al.*, 2013). De hecho, se ha demostrado recientemente que en el caso de las DNAPs eucariotas Pol  $\alpha$ , Pol  $\delta$  y la Pol  $\epsilon$ , con una tasa de error *in vitro* en el rango de  $10^{-4}$ - $10^{-5}$  (Kunkel *et al.*, 1991, Shcherbakova *et al.*, 2003, Fortune *et al.*, 2005), su fidelidad se ve incrementada hasta  $10^{-7}$  *in vivo* (St Charles *et al.*, 2015). Este cambio podría deberse a las diferencias en las concentraciones de sal o en el pH entre ambos escenarios o a algún otro mecanismo que incremente la fidelidad *in vivo*.

Hemos determinado que el error más común producido por la B35DNAP son las inserciones en secuencias moldes homopoliméricas siendo las deleciones y substituciones muy poco frecuentes. El mismo análisis en otras DNAPs (Pol  $\gamma$ ,  $\delta$ ,  $\epsilon$ ...) muestran que la tasa de error por substituciones es mucho más alto que por inserciones o deleciones (McCulloch y Kunkel, 2008), a excepción de la Pol  $\theta$  humana que pertenece a la familia A. La Pol  $\theta$  presenta un rango de error de  $10^{-3}$ - $10^{-4}$  que es mayor que las DNAPs replicativas y muestra, principalmente, inserciones de nucleótidos ( $3,3 \times 10^{-3}$ ) y deleciones ( $1,4 \times 10^{-3}$ ), sobre todo en moldes con secuencias homopoliméricas (Maga *et al.*, 2002, Arana *et al.*, 2008). Las inserciones y deleciones se producen por un error en el apareamiento entre el *primer* y el molde, conteniendo un nucleótido extra en la cadena molde con respecto al producto en el caso de deleciones o en el *primer* en el caso de las inserciones. Este error se estabiliza cuando existen nucleótidos correctamente apareados entre el extremo del *primer* que se encuentra en el centro catalítico de la DNAP y el nucleótido extra (García-Díaz y Kunkel, 2006). Por lo tanto, la frecuencia mayor de inserciones con respecto a las deleciones observada en la replicación por la B35DNAP sugiere que los contactos establecidos entre esta enzima y la cadena molde del DNA serían muy estables y evitarían la pérdida de información genética al favorecer la copia de todos los nucleótidos. Sin embargo, la interacción de la B35DNAP con el DNA podría permitir el desapareamiento del *primer* resultando en la copia reiterativa del mismo nucleótido molde. Este tipo de error fue más frecuente en secuencias homopoliméricas, sugiriendo que este tipo de secuencias pueden producir una acumulación de mutaciones que cambien el marco de lectura.



## 2. Iniciación de la replicación del genoma de Bam35 mediada por TP

En esta tesis hemos demostrado que, tal y como se había sugerido previamente (Sozhamannan *et al.*, 2008, Redrejo-Rodríguez *et al.*, 2012, Jalasvuori *et al.*, 2013), el ORF4 de Bam35 codifica la TP utilizada como *primer* en la replicación de su genoma.

### 2.1. Deoxinucleotidilación de la TP por la B35DNAP

La B35DNAP es capaz de formar un enlace fosfoéster entre la TP y un dAMP en ausencia de molde y

con ambos metales divalentes ensayados (manganeso y magnesio), aunque, al igual que ocurre en otros sistemas como PRD1 o  $\Phi 29$  (Caldentey *et al.*, 1992, Esteban *et al.*, 1992), la reacción es mucho más eficiente con manganeso. Además, el pH también afecta a la eficiencia de reacción, sobre todo cuando el magnesio es el metal utilizado, de tal manera que no se detecta ningún producto de reacción hasta alcanzar pHs básicos (8-9), de acuerdo con el alto punto isoeléctrico de la TP (alrededor de 10,6). Sin embargo, el efecto del pH es menor con manganeso. Además, la B35DNAP muestra una gran preferencia por dATP para llevar a cabo la iniciación en ausencia de molde, seguida de dCTP, mientras que con dGTP o dTTP no se observa ningún producto. Este resultado es similar al descrito para la DNAP de PRD1 la cual también inserta preferentemente dAMP, pero en este caso la eficiencia con dCTP, dGTP y dTTP es prácticamente la misma, dando lugar con todos ellos a una pequeña banda correspondiente al producto de reacción (Caldentey *et al.*, 1992). Por otro lado, la preferencia de B35DNAP difiere de la  $\Phi 29$ DNAP en que el producto obtenido con dATP, dGTP y dTTP es prácticamente el mismo, mostrando una eficiencia algo menor con dCTP (Blanco *et al.*, 1992), y de adenovirus en el que el uso de dCTP es el preferido (King y van der Vliet, 1994). Esto sugiere que la preferencia por dATP puede ser una característica común entre los tectivirus. Aunque el posible papel biológico de esta reacción no está claro puesto que los nucleótidos unidos a la TP en ausencia de molde difieren en la base necesaria para la iniciación de la replicación, timina en el caso de Bam35 o guanina en PRD1. Por otro lado, no podemos descartar que la elevada eficiencia de deoxiadenilación de la TP en ausencia de molde producida por la B35DNAP pudiera ser necesaria para otros fines no directamente relacionados con la replicación.

## 2.2. La replicación del tamaño completo del TP-DNA *in vitro* requiere solo de dos proteínas

Al igual que la  $\Phi 29$ DNAP y la DNAP de PRD1, la B35DNAP es capaz de replicar *in vitro* el tamaño completo de su genoma añadiendo únicamente la TP como proteína adicional. Sin embargo, la reacción es menos eficiente que las descritas para las proteínas de  $\Phi 29$  y PRD1 (Blanco *et al.*, 1989, Caldentey *et al.*, 1992) necesitando tiempos muy largos (2 horas) para obtener el producto final. En el caso de  $\Phi 29$ , la presencia de proteínas accesorias como la DBP y SSB incrementan la eficiencia de la replicación del genoma *in vitro* mientras que en otros sistemas como el de adenovirus o Nf, la presencia de la DBP es absolutamente necesaria (de Jong *et al.*, 2003, Longás *et al.*, 2006). La baja eficiencia observada en el caso de Bam35 sugiere la necesidad de otras proteínas accesorias para incrementar su replicación. El hecho de que, aunque la iniciación observada con B35DNAP sea más eficiente con manganeso que con magnesio, no se obtenga un producto de replicación del tamaño del genoma con manganeso, podría ser explicado por una pérdida en la procesividad de la B35DNAP al utilizar este metal, del mismo modo que ocurre en el caso de PRD1 (Caldentey *et al.*, 1992) y  $\Phi 29$  (Esteban *et al.*, 1992, Méndez *et al.*, 1992), en los cuales la presencia de magnesio es esencial para la replicación. Así, el uso de esos cofactores metálicos por las DNAPs es controvertido y afecta fuertemente a

su funcionamiento y fidelidad (Beckman *et al.*, 1985, Esteban *et al.*, 1993). Además, no hay que olvidar que su disponibilidad *in vivo*, sobre todo la del manganeso, está altamente regulada por la bacteria (Moore y Helmann, 2005).

### 2.3. Los residuos conservados tirosina 194 y 172 de la TP son esenciales para la iniciación

El residuo de la TP utilizado por la B35DNAP para iniciar la replicación es la tirosina 194, que está conservada entre bacteriófagos de la misma familia como Gil16 o AP50. Además, al hacer un alineamiento entre TPs relacionadas podemos observar secuencias anotadas como pertenecientes a bacterias como *B. cereus*, *S. pneumoniae*, *Exiguobacterium*, *B. flexus* y *Brevibacillus* sp. en las que también está presente esa tirosina, y que, debido a la presencia adicional de DNAPs de la familia B y otras proteínas de fagos, probablemente, codifiquen TPs pertenecientes a profagos o plásmidos similares a tectivirus que aún no han sido caracterizados. Recientemente se ha propuesto la existencia de una relación entre la posición del extremo 3' del genoma utilizada como molde para la iniciación de la replicación y la posición de un residuo aromático N-terminal con respecto al residuo iniciador de la TP (Del Prado *et al.*, 2015). De esta manera, se ha observado que en el caso de  $\Phi$ 29 y GA-1, que inician la replicación utilizando la segunda base del molde, sus residuos iniciadores serina 232 y serina 231, respectivamente, tienen una fenilalanina localizada a dos posiciones (**Figura 49**). Esta relación también se comparte con Cp-1 que inicia frente a la tercera posición del extremo 3' y su residuo aromático (fenilalanina 186) está a 3 posiciones del residuo iniciador (treonina 189). En el caso de adenovirus, que inicia frente a la cuarta posición del extremo 3' de la cadena molde, el residuo aromático (tirosina 576) está situado a 4 posiciones del residuo iniciador (serina 580). Y, en el caso de PRD1, su residuo aromático (tirosina 187) se encuentra a 3 posiciones del residuo iniciador (tirosina 190) y la iniciación se produce frente a la cuarta base. Curiosamente, en el caso de Bam35, no se

|                            |       |                    |
|----------------------------|-------|--------------------|
| <b><math>\Phi</math>29</b> | /226/ | SFEE <b>FD</b> SEG |
| <b>GA-1</b>                | /225/ | SFEN <b>FD</b> SEN |
| <b>Cp-1</b>                | /183/ | VMV <b>FVD</b> TDG |
| <b>Adenovirus</b>          | /574/ | IA <b>YQDN</b> SGD |
| <b>PRD-1</b>               | /184/ | FNK <b>FMT</b> YDI |
| <b>Bam35</b>               | /188/ | TIMVGV <b>Y</b> DV |

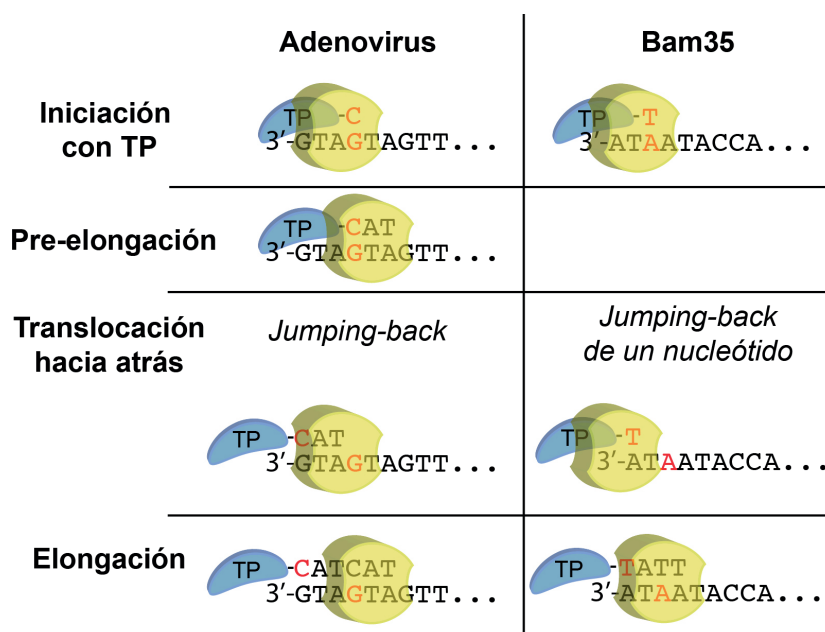
**Figura 49. Secuencia de aminoácidos del bucle iniciador de las TPs de diferentes virus.** Los números entre las barras indican la posición del primer aminoácido indicado con respecto al extremo N-terminal de la TP. En rojo está indicado el residuo iniciador y en verde el aminoácido aromático cercano a ese residuo. Figura modificada de Del Prado *et al.* (2015)

observa ningún residuo aromático más cerca de 22 posiciones (corresponde a la tirosina 172) con respecto al residuo iniciador (tirosina 194). Estas diferencias observadas entre Bam35, y los fagos  $\Phi 29$ , GA-1, Cp-1, PRD1 y adenovirus, sugieren que, en el caso de Bam35, la determinación de la posición de inicio podría estar regulada por un mecanismo diferente.

La tirosina 172 también está conservada entre las secuencias de las TPs relacionadas y parece que está implicada en la interacción con la DNAP. Esta hipótesis proviene del hecho de que cuando la tirosina 172 es cambiada por una alanina, la TP no puede ser deoxiadenilada por la B35DNAP mientras que, por el contrario, cuando se modifica por una fenilalanina, la deoxiadenilación se ve incrementada. Este incremento en la deoxiadenilación podría deberse a una interacción reforzada entre la TP Y172F y la B35DNAP con respecto a la TP *wild type*. Puesto que la única diferencia entre la fenilalanina y la tirosina es la ausencia o presencia de un grupo hidroxilo, podemos sugerir que la tirosina 172 está implicada en una interacción de tipo hidrofóbica con la B35DNAP.

#### 2.4. Un nuevo mecanismo de recuperación de la información genética del extremo 3' del genoma: *Jumping-back* de un nucleótido

Los resultados obtenidos en la replicación truncada de oligonucleótidos sugieren fuertemente que, después del inicio de la replicación utilizando la TP como *primer*, el complejo B35DNAP/TP-dTMP se transloca desde la tercera posición de la cadena molde hasta la primera posición para recuperar la información genética de los dos primeros nucleótidos de los extremos 3' mediante un mecanismo al que denominamos *jumping-back* de un nucleótido (**Figura 50**). Este mecanismo se diferencia del



**Figura 50. Translocación hacia atrás mediante *jumping-back*.** El esquema representa el *jumping-back*, previamente descrito, de adenovirus, y el *jumping-back* de un nucleótido sugerido en el caso de Bam35 y descrito por primera vez en esta tesis. La DNAP está representada en color amarillo.

*jumping-back* descrito en el caso de adenovirus en el que la DNAP copia tres nucleótidos antes de la translocación (**Figura 50**) (King y van der Vliet, 1994). Si tenemos en cuenta la secuencia de ambos orígenes del genoma de Bam35, 3'-ATAATA..., se puede sugerir que este mecanismo podría ser el favorecido en la replicación del genoma de Bam35 *in vivo*. Por el contrario, PRD1 que infecta a bacterias Gram-negativas utiliza un mecanismo de *stepwise sliding-back* (Caldentey *et al.*, 1993) para recuperar la información genética. La similitud del mecanismo del *jumping-back* utilizado por Bam35 y adenovirus junto al análisis filogenético que sugiere que unos fagos similares a los tectivirus podrían ser el origen de algunos virus eucariotas de dsDNA (Introducción apartado 2, página 15), sugieren que el mecanismo de replicación de Bam35 podría estar en el origen del sistema de replicación de esos virus.

Al analizar la secuencia de los extremos del genoma de fagos cercanamente relacionados a Bam35, se observa que en algunos casos como Gil01 y Gil16 solo difieren en la primera base de uno o ambos extremos (**Figura 51**). Es posible que estas variaciones se deban a las dificultades técnicas para secuenciar los extremos del DNA que contienen TP, por lo que podemos especular con que esos tres fagos repliquen sus genomas utilizando el mismo mecanismo. Sin embargo, en el caso de AP50 no se observan ITRs que permitan un mecanismo de translocación reversa (**Figura 51**). Por tanto, para poder determinar si el mecanismo de *jumping-back* de un nucleótido es común entre los tectivirus que infectan a Gram positivas y poder realizar estudios evolutivos, será necesario tener disponibles más secuencias de sus extremos del genoma y estudiar sus mecanismos de replicación.

| Bacteriófago | Extremo del genoma                                |
|--------------|---|
| <b>Bam35</b> | 5'-TATTATGGTACCCCTACCAA...CTAGTAGGGGTACATAATA-3'  |
| <b>Gil01</b> | 5'-TATTATGTACCCCTACCAAC...GCTAGTAGGGGTAATAATA-3'  |
| <b>Gil16</b> | 5'-TATTATGTACCTCTACTAGC...GCTAGTAGGGGATAATAATA-3' |
| <b>AP50</b>  | 5'-CATGGTGTTCAGCCCATCTA...CAGATGGGCTGAACACCATG-3' |

**Figura 51. ITRs de los extremos de los genomas de los tectivirus Bam35, Gil01, Gil16 y AP50.** En naranja están marcados los nucleótidos conservados entre Bam35, Gil01 y Gil16, y en naranja claro el posible primer nucleótido no detectado por secuenciación de Gil01 y Gil16. La secuencia de los extremos de AP50 no muestra homología con los otros bacteriófagos representados.



### 3. La B35DNAP: Una enzima con alta fidelidad capaz de replicar DNAs con sitios abásicos

Los sitios abásicos son un tipo de daño en el DNA que bloquea la replicación de las DNAPs replicativas. Diversos estudios han demostrado que este tipo de DNAPs muestra una capacidad muy baja para replicar a través de estos daños, siendo habitualmente dependiente del contexto de secuencia y de la presencia o ausencia de la actividad exonucleasa 3'-5'. Así, se han descrito variantes de estas DNAPs con capacidad



para replicar con éxito un DNA con sitios abásicos, siendo una de las variaciones más determinantes la eliminación de la actividad correctora de pruebas 3'-5' (Tanguy Le Gac *et al.*, 2004, Zhu *et al.*, 2008, Jozwiakowski *et al.*, 2014).

A pesar de la baja tasa de error mostrada ( $5,1 \pm 1,7 \times 10^{-6}$ ) y de ser una DNAP replicativa, la B35DNAP *wild type* es capaz de replicar oligonucleótidos con THFs (análogo estable de un sitio abásico) en su secuencia de una manera muy eficaz. Durante el estudio de la replicación de este tipo de molde, siempre se observó un producto de 16-mer que correspondía al *primer* al cual la B35DNAP insertó un único nucleótido, generando un no apareamiento entre ese nucleótido y el THF situado enfrente. La extensión de este no apareamiento es el paso limitante en la replicación de DNAs con sitios abásicos debido a la capacidad de la B35DNAP para detectarlo y eliminarlo mediante su actividad exonucleasa (ver página siguiente). La presencia de este producto sugiere que algunas de las moléculas de B35DNAP que no son capaces de extender ese error, se disocian del oligonucleótido abortando la replicación. Sin embargo, hemos comprobado que los *primers* elongados con éxito son el resultado de una replicación procesiva llevada a cabo por una única molécula de B35DNAP que no se disocia del DNA molde. Además, la obtención de un producto final del tamaño completo de la cadena molde (33-mer) sugiere que la B35DNAP introduce un nucleótido enfrente del THF en lugar de generar un re-alineamiento entre el *primer* y el nucleótido siguiente de la cadena molde para superar ese daño, lo que tendría como consecuencia un producto de 1 o 2 nucleótidos menos que el oligonucleótido molde, tal como sucede en el caso de la Pol  $\kappa$  o la Pol II (Ohashi *et al.*, 2000, Wang y Yang, 2009). Este resultado está en concordancia con el obtenido en el análisis de la fidelidad de la B35DNAP que indica que esta enzima tiene escasa tendencia a producir deleciones. Además, el análisis de la polimerización, paso a paso, del oligonucleótido dañado por la variante sin actividad exonucleasa de la B35DNAP, confirma que esta enzima inserta un nucleótido enfrente del sitio abásico. Finalmente, la preferencia de inserción de dAMP enfrente del THF, independientemente del nucleótido molde que hay a continuación, confirma que la inserción del nucleótido ocurre únicamente enfrente del daño. Esta preferencia se ha observado en más DNAPs y se conoce con el nombre de la “regla de la A”, y tiene como consecuencia la acumulación de mutaciones en la nueva cadena de DNA sintetizada (Strauss, 2002). En conjunto, todos estos resultados indican que la B35DNAP introduce una A enfrente del sitio abásico antes de la elongación del *primer* y descartan la posibilidad de que la B35DNAP produzca un re-alineamiento entre el *primer* y el nucleótido siguiente de la hebra molde para superar ese daño.

A pesar de que el contexto de secuencia no afecta al nucleótido insertado enfrente del THF por la B35DNAP, si influye en la capacidad para insertarlo y en su posterior elongación, al igual que se ha descrito para otras DNAPs como la DNAP humana  $\eta$ , la Dpo4 de *S. solfataricus* y la DNAP de levaduras Pol  $\epsilon$  (Kokoska *et al.*, 2003, Sabouri y Johansson, 2009). No obstante, la B35DNAP tiene una eficiencia de replicación a través

de THF más alta que otras DNAPs replicativas como la Pol  $\epsilon$  de levaduras (Sabouri y Johansson, 2009) y comparable a algunas DNAPs de la familia Y como la DNAP humana  $\eta$ , la DNAP IV de *E. coli* y la Dpo4 de *S. solfataricus* (Kokoska *et al.*, 2003, Maor-Shoshani *et al.*, 2003).

### 3.1. La “Memoria a corto plazo” de la capacidad correctora de pruebas contrarresta la replicación de DNA dañado

La B35DNAP es capaz de detectar el error de apareamiento entre el THF y el nucleótido insertado enfrente hasta que se encuentra a 5 o 6 posiciones *upstream* del extremo 3' del *primer*. Estudios previos han demostrado que la capacidad de las DNAPs replicativas para elongar el *primer* se ve influenciada por la presencia de daños o errores de apareamiento en el DNA cuando se encuentran a 1-5 posiciones *upstream* del extremo 3' del *primer* (Ng *et al.*, 1989, Miller y Grollman, 1997, Johnson y Beese, 2004). Cuando esos errores se localizan entre dichas posiciones, se produce la distorsión del DNA, principalmente de su hebra molde, en el centro activo de la DNAP produciendo la parada de la replicación debido a la denominada “memoria a corto plazo” (Johnson y Beese, 2004). De esta manera, las bases en las posiciones 3 y 4 *upstream* del molde se trasladan al surco menor de la DNAP en una conformación que es estabilizada por el contacto directo entre la enzima y el esqueleto del DNA y los contactos mediados por los disolventes del surco menor, de manera que si esta estabilización no es correcta, se produce la parada de la replicación. En el caso de la  $\Phi$ 29DNAP, las interacciones entre el surco menor y el DNA se estabilizan a través de unos residuos conservados en el subdominio *thumb* que interaccionan con los nucleótidos o el esqueleto azúcar-fosfato del *primer* hasta 5 posiciones *upstream* (Pérez-Arnaiz *et al.*, 2006, Berman *et al.*, 2007). Así, del mismo modo, el subdominio *thumb* de la B35DNAP podría interaccionar con el *primer*, permitiendo detectar el error en el apareamiento durante una vuelta de hélice del DNA, proporcionando la “memoria” durante su polimerización. Probablemente esta “memoria” también permita detectar un nucleótido introducido de manera errónea hasta que se encuentre a 5 posiciones *upstream* del extremo 3' del *primer* (Johnson y Beese, 2004). Parece que, en el caso de los sitios abásicos, esta “memoria a corto plazo” se reduce durante la síntesis procesiva del DNA dando lugar a moléculas del mismo tamaño que el molde en presencia de los cuatro dNTPs, sugiriendo que el error puede “escapar” a la actividad exonucleasa durante la replicación procesiva del genoma, logrando así un equilibrio entre replicación procesiva y fiel y la capacidad de síntesis a través de lesiones. Esto explicaría por qué, al eliminar el subdominio TPR2 de la B35DNAP implicado en la procesividad, la B35DNAP pierde la capacidad de elongar el *primer* y únicamente es capaz de introducir un nucleótido enfrente del THF.

Aunque la replicación de DNA con sitios abásicos por la B35DNAP produzca la acumulación de mutaciones, su replicación es importante para permitir la integridad del genoma. Esta importancia se ve remarcada

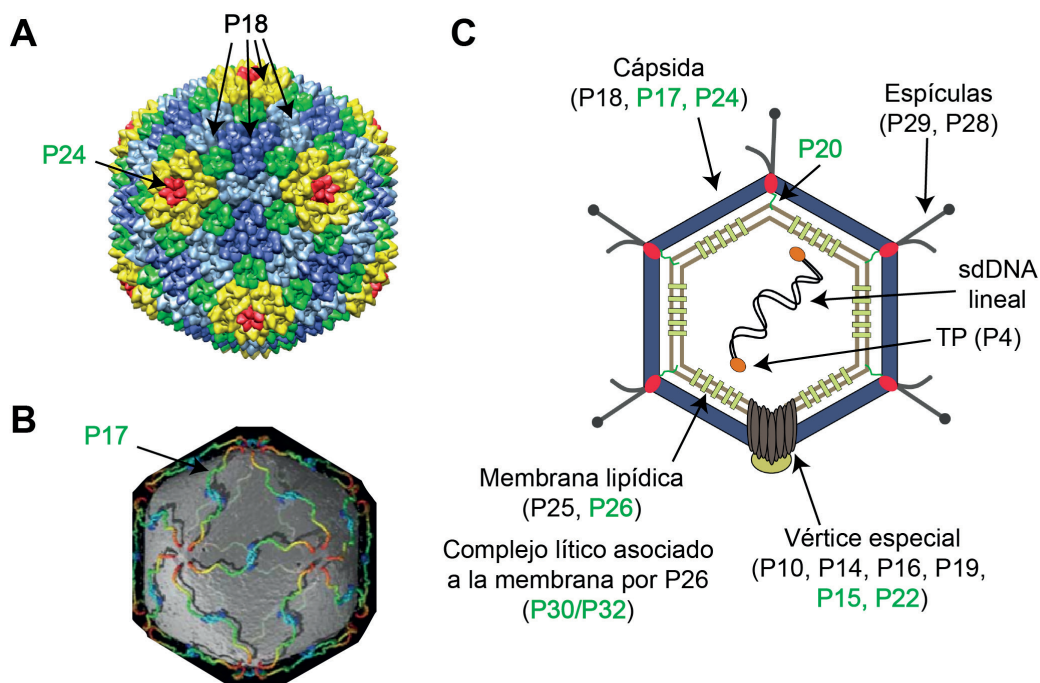


por el hecho de que Bam35 y otros bacteriófagos relacionados presentan un ciclo de vida lisogénico lo que les condiciona a auto-replicarse durante generaciones dentro de la bacteria hospedadora al mismo tiempo que su genoma puede verse expuesto, junto al genoma del hospedador, a agentes genotóxicos y exógenos que pueden producir daños en el DNA, entre los que destaca la generación de sitios abásicos (Lindahl, 1993). Este efecto podría ser de especial importancia en el hospedador de Bam35, *B. thuringiensis*, el cual muestra una mayor sensibilidad a daños en el DNA por agentes alquilantes que otras especies de *Bacillus* (Morhoshi y Munakata, 1995). Como consecuencia de todo lo expuesto anteriormente, podría sugerirse que la B35DNAP ha evolucionado para salvaguardar la integridad del genoma viral *in vivo* mediante una replicación procesiva con alta fidelidad acoplada a la replicación de ciertos daños en el DNA cuando fuese requerido.



#### 4. La red de interacción de las proteínas de Bam35 revela el papel de nuevas proteínas estructurales

Mediante el análisis de la interacción entre proteínas se puede especular la función o la localización de proteínas desconocidas. Con este fin, se realizó dicho análisis entre todas las proteínas codificadas por el genoma de Bam35 mediante el sistema de doble híbrido de levaduras. De esta manera, integrando los resultados obtenidos en esta tesis con los datos bibliográficos previos, podemos proponer alguna de las



**Figura 52. Modelo propuesto de la estructura de la partícula viral del bacteriófago Bam35.** Las proteínas propuestas a través del estudio de las interacciones obtenidas mediante el sistema de doble híbrido de levaduras se indican en verde. A) Morfología de la cápsida de PRD1 obtenida mediante cristalografía de rayos X (PDB 1W8X) utilizada como modelo para Bam35. La proteína mayoritaria de la cápsida P18 de Bam35 se indica en color amarillo, azul, azul claro y verde. La proteína propuesta como formadora del pentón se indica en rojo. B) Proteína propuesta en esta tesis como minoritaria de la cápsida. C) Esquema del virión de Bam35 con las proteínas conocidas o propuestas para cada estructura.

proteínas que conforman la partícula viral de Bam35 (**Figura 52**). Sin embargo, no hay que olvidar que el sistema de doble híbrido puede generar en torno a un 70-80% de falsos negativos, aunque este porcentaje se ve reducido hasta un 20% al utilizar varias combinaciones de vectores durante el análisis (Stellberger *et al.*, 2010).

#### 4.1. Proteínas de la cápsida y de la espícula

Lo primero que llama la atención al analizar el interactoma proteico (**Figura 48**) es que la proteína mayoritaria de la cápsida (P18) (Ravantti *et al.*, 2003), además de interactuar consigo misma como era de esperar, únicamente interactúa con la P17 (9,7 KDa). La interacción consigo misma confirma la utilidad del sistema de doble híbrido para detectar interacciones entre proteínas estructurales virales. Por otro lado, la interacción con la P17 sugiere que esta proteína puede ser la minoritaria de la cápsida, teniendo un papel similar a la P30 en PRD1 (9,2 KDa) (Abrescia *et al.*, 2004). No obstante, no podemos descartar completamente que se encargue de formar el pentón (P31 en PRD1, 13,7 KDa), debido a su interacción con la proteína de la espícula P28 (homóloga a la P5 de PRD1 (Ravantti *et al.*, 2003)) y con la proteína P29, que se ha sugerido que actúa como proteína de unión al receptor del hospedador (Laurinmaki *et al.*, 2005). La proteína P29 (32,7 KDa) tiene la mitad de peso molecular que su homóloga en función en PRD1 (P2, 63,8 KDa) y, mientras que, en PRD1 aparece como un monómero, en Bam35 podría formar complejos al interactuar consigo misma generando así diferencias en la morfología de la espícula. Por otro lado, la proteína P24 ha sido propuesta como proteína minoritaria de la cápsida (Sozhamannan *et al.*, 2008), aunque nuestros resultados no apoyan esa hipótesis debido a la ausencia de interacción con la proteína mayoritaria P18. Sin embargo, su peso molecular, 10,7 KDa, y la interacción consigo misma y con la proteína P29 observada en nuestros resultados sugieren que la proteína P24 es la que conforma el pentón. Aunque previamente se había propuesto la proteína P13 como posible proteína del pentón (Laurinmaki *et al.*, 2005), nuestros resultados descartan esta función porque, aunque la P13 interactúa con la P29, no se observa con ninguna de las fusiones expresadas interacción de la P13 consigo misma, requisito indispensable para formar pentámeros. Finalmente, y como era de esperar, la proteína de la espícula P28 interactúa consigo misma y con la P29. Sin embargo, contrariamente a lo esperado, la proteína P28 no interactúa con la posible proteína del pentón P24 por lo que sería necesario realizar estudios adicionales de interacción entre ambas proteínas para confirmar el posible papel de la proteína P24.

Al analizar la interacción entre las proteínas que forman el vértice por el que se empaqueta el DNA en la partícula viral, encontramos que la proteína de empaquetamiento de DNA P19 (Stromsten *et al.*, 2003a) interactúa con la proteína P22 que a su vez interactúa con la proteína de membrana P15, lo que localizaría a ambas proteínas cerca del vértice especial. La localización de la P22 en este vértice está en

concordancia con la función en la liberación del DNA propuesta para su ortólogo en AP50 (Sozhamannan *et al.*, 2008). Por otro lado, la proteína P14 localizada en ese vértice (Stromsten *et al.*, 2003a) está relacionada con la P22 a través de la interacción con la proteína P26 que tiene actividad lítica (Gaidelyte *et al.*, 2006).

#### 4.2. Proteínas de membrana

Al realizar los perfiles de hidrofobicidad de las proteínas de Bam35, se detectaron 9 proteínas con alta probabilidad de poseer dominios de transmembrana; este resultado difiere de las 10 proteínas descritas originalmente por Ravantti *et al.* (2003). En ese trabajo se indica que la proteína P32 también poseería este tipo de dominio. Sin embargo, al analizar este dato (Tabla 2 del artículo) nos damos cuenta de que probablemente se deba a un error, puesto que la longitud en pares de bases del ORF32 son 309 y no 425, error que ha sido arrastrado a lo largo de la bibliografía.

Como era de esperar, los resultados obtenidos con las proteínas sin el dominio de transmembrana difirieron de los obtenidos con las proteínas completas. Existen diversas causas para estas diferencias, por ejemplo, que las proteínas completas no se hayan expresado o lo hayan hecho formando agregados y/o que la estructura terciaria de las proteínas modificadas sean diferentes a las de las proteínas completas y por lo tanto presenten en su superficie dominios internos generando falsos positivos, y/o que el dominio eliminado sea necesario para interactuar con otras proteínas. Teniendo en cuenta estas posibilidades podemos sugerir que la ausencia total de interacciones con las proteínas P19' y P21', en contraste con la presencia con las versiones completas podría deberse a que el dominio eliminado en esas proteínas fuese necesario para estas interacciones. En el caso de la versión corta de la proteína mayoritaria de la membrana (P25') coincidió en la interacción con las proteínas estructurales P17, P24 y P29 con su versión completa (P25), dando credibilidad a estas interacciones. Sin embargo, la P25' mostró alta promiscuidad cuando fue expresada a partir del vector pGADT7g (N) y un nivel de auto-activación muy alto desde el vector pGADCg (C), lo que nos llevó a catalogar las otras interacciones observadas como falsos positivos (**Figura 42**). Por otro lado, la proteína P20 no mostró ninguna interacción mientras que la P20' interactuó con la proteína de función desconocida P3 y las estructurales P24 y P29. Sin embargo, estas interacciones fueron muy débiles, observándose tan solo en una combinación de vectores y en ausencia de 3AT. No obstante, por las interacciones mostradas, podemos sugerir que la P20 podría ser la proteína de transmembrana encargada de estabilizar las espículas teniendo una función homóloga a la P16 de PRD1. Sin embargo, debido a la debilidad de las interacciones son necesarios estudios adicionales para corroborar este resultado. Por otro lado, la proteína de membrana con actividad transglicosilasa P26 dio lugar a diferentes interacciones según la construcción ensayada. Así, la proteína P26 interactuó con las proteínas P21 y P22, las cuales no fueron detectadas con la P26', sugiriendo que el dominio eliminado (aminoácidos 117-250) es necesario

para interactuar con esas proteínas. Por su parte, la P26' interactuó con la P17, P24, P29, P30, P32 y consigo misma de un manera muy fuerte dando mucha fiabilidad al resultado. Además, la interacción con la proteína de la cápsida P29 está en línea con trabajos previos en los que se ha descrito que el virión de Bam35 posee proteínas con capacidad de hidrolizar peptidoglicanos (Gaidelyte *et al.*, 2006). De este mismo modo, la interacción de la proteína P26' con la transglicosilasa P30 y, a su vez, la interacción de la proteína P30 con las proteínas P17, P24 y P29 localizan a esta proteína también en el virión. Todas estas interacciones, sugieren fuertemente, que además de la función lítica demostrada previamente para la P26, esta proteína estaría implicada en más funciones, bien durante la morfogénesis de la partícula o como proteína de anclaje de otros elementos estructurales.

Cabe destacar que, mientras que la proteína P31 formada por los aminoácidos 50 al 90 de la proteína P30, no dio lugar a ninguna interacción, la proteína P32 formada por la parte C-terminal (aminoácidos 162-265) coincidió en todas las interacciones con la proteína P30 a excepción de la interacción con la P29, lo que indicaría que la proteína P30 interactúa con esa proteína por su lado N-terminal y más concretamente con los aminoácidos 1-162. Además, las proteínas P30 y P32 interactúan entre sí, recordando al complejo formado por las proteínas P7 y P14 descrito recientemente en el virión de PRD1 (Mattila *et al.*, 2015). La proteína P14 de PRD1 corresponde con la parte C-terminal de la proteína P7, al igual que ocurre con la proteína P32 y P30 de Bam35 y, en ambos casos, las proteínas largas (P7 de PRD1 y P30 de Bam35) tienen actividad transglicosilasa proporcionada por el dominio N-terminal mientras que las proteínas cortas (P14 de PRD1 y P31 de Bam35) corresponden con la otra parte de la proteína. Sin embargo, ambas proteínas de PRD1 poseen un pequeño dominio de transmembrana que les permite anclarse a la membrana, dominio que es inexistente en las proteínas de Bam35. Por lo tanto, las similitudes observadas sugieren que las proteínas P30 y P32 forman un complejo similar al descrito en PRD1 y que probablemente estén asociadas a la membrana a través de la interacción con otra proteína de anclaje. Nuestros resultados sugieren que las proteínas P30 y P32 únicamente interactúan con una proteína de membrana, la proteína P26', pudiendo ser dicha proteína el anclaje a la membrana utilizado por ese complejo.

### 4.3. Proteínas de replicación y regulación del ciclo

Otro de los objetivos del análisis con el sistema de doble híbrido fue encontrar proteínas que interactuasen con la B35DNAP (P5) y/o con la TP (P4) y que por tanto pudiesen estar implicadas en la replicación del genoma de Bam35. En esta tesis se ha demostrado que la B35DNAP y la TP interactúan entre sí mediante ensayos funcionales (Capítulo 2 de Resultados). Sin embargo esta interacción no puede ser detectada mediante otras técnicas como el gradiente de glicerol probablemente debido a que la interacción no es lo suficientemente estable para resistir esas condiciones (datos no mostrados). Esta inestabilidad podría

explicar porque dicha interacción tampoco es detectable mediante el sistema de doble híbrido. Sin embargo, no podemos descartar que la ausencia de crecimiento de la levadura que expresa ambas proteínas sea consecuencia del alto punto isoeléctrico de la TP que le proporciona una alta afinidad por el DNA, al igual que ocurre con la TP de  $\Phi 29$  (Zaballos y Salas, 1989) lo que podría dificultar tanto la interacción con otras proteínas como el crecimiento de la levadura. A pesar de esto, se observó que la TP interaccionaba consigo misma. Este resultado sugiere la posibilidad de la interacción de la TP iniciadora con la TP parental implicándola en el reconocimiento del origen de replicación del genoma de Bam35. Esta posibilidad también se ha propuesto para la TP de  $\Phi 29$  puesto que se ha observado que diversos mutantes de esta proteína están afectados en el reconocimiento del origen (Illana *et al.*, 1998, Serna-Rico *et al.*, 2000, Del Prado *et al.*, 2012), aunque la DNAP también contribuye en este aspecto (González-Huici *et al.*, 2000, Pérez-Arnaiz *et al.*, 2007). Además, la interacción entre TPs también está apoyada por el hecho de que el TP-DNA lineal de  $\Phi 29$  es capaz de circularizarse y formar concatémeros gracias a la interacción entre las TPs parentales (Ortín *et al.*, 1971, Salas *et al.*, 1978).

La TP de Bam35 (P4) también interaccionó con la proteína P1. La P1 tiene un dominio de unión al DNA similar al encontrado en la superfamilia de represores de operones de resistencia a mercurio de los reguladores transcripcionales (Fornelos *et al.*, 2011). Esta superfamilia media la activación o represión de la transcripción mediante su capacidad para distorsionar la hélice del DNA. Por eso, previamente, se ha sugerido que la P1 podría tener algún tipo de función asociada a la modulación de la estructura del DNA (Fornelos *et al.*, 2011). Esto, junto a los resultados recientemente obtenidos en nuestro laboratorio en los que se demuestra que la P1 se une a DNA *in vitro* (resultados no publicados) y a la interacción encontrada entre la TP y la P1, sugieren que la interacción con la TP podría atraer a la P1 al origen de replicación del TP-DNA para facilitar la iniciación y la posterior replicación.

Con respecto a las proteínas de regulación del ciclo viral de Bam35, encontramos que la proteína P6, inicialmente propuesta como candidata para actuar como un represor del ciclo lítico (Fornelos *et al.*, 2011), interaccionó con un gran número de proteínas entre las que destacan la P29, P24 y P30. Sin embargo, son necesarios estudios adicionales para encontrar una explicación biológica para estas interacciones. Por otro lado, la proteína P7, que actúa, a través de la interacción con el represor LexA del hospedador, como regulador del ciclo lisogénico de Bam35 (Fornelos *et al.*, 2015), mostró una alta capacidad para interaccionar consigo misma. Esto podría indicar la necesidad de formar complejos para regular el ciclo del bacteriófago, ya sea porque los complejos de la proteína P7 tengan una mayor afinidad por el represor LexA que su forma monomérica o porque la formación de complejos impidan la unión de P7 al represor. Al igual que la proteína P6, la P7 también interaccionó, de una manera débil, con las proteínas P24 y P26' necesitando, de nuevo, más estudios para determinar el papel biológico de estas interacciones.



## 5. Conclusión: relevancia de los resultados y perspectivas futuras

En esta tesis se ha establecido que la B35DNAP es una enzima que presenta una replicación procesiva de alta fidelidad acoplada al desplazamiento de banda. Estas características, junto a su capacidad para replicar un DNA con sitios abásicos sin modificar el marco de lectura, podrían permitir el empleo de la B35DNAP para la amplificación *in vitro* de DNA dañado. Dicho uso podría ser muy útil para aplicaciones paleontológicas o forenses.

Por otro lado, los resultados obtenidos en esta tesis han establecido las bases de un nuevo modelo de estudio de replicación del DNA y han permitido el planteamiento de nuevas cuestiones como la búsqueda de proteínas adicionales implicadas en la replicación del genoma de Bam35, las cuales se están intentando resolver en nuestro laboratorio. Además de un interés científico intrínseco, la optimización del sistema de replicación de Bam35 permitiría el desarrollo de posibles aplicaciones biotecnológicas del mismo. Así por ejemplo, la capacidad de la B35DNAP para llevar a cabo la replicación del DNA utilizando una proteína terminal como *primer* podría ser aprovechada, al igual que en el caso de adenovirus y  $\Phi 29$ , para generar DNAs unidos a proteínas (Mencia *et al.*, 2011, Redrejo-Rodríguez *et al.*, 2013b, Holkers *et al.*, 2014). En este sentido, el ciclo lisogénico de Bam35 presentaría una ventaja interesante respecto a  $\Phi 29$  y adenovirus puesto que permitiría que el genoma del fago, en el cuál podremos haber insertado un gen de interés, se transmita de generación en generación dentro de la bacteria hospedadora.

Finalmente, el sistema de doble híbrido en levaduras nos ha permitido sugerir la función de algunas de las proteínas de función desconocida codificadas por el genoma del bacteriófago Bam35 a través del estudio de sus interacciones. Estos resultados nos proporcionan un mayor entendimiento de la biología del fago y sientan las bases para la elaboración de futuros proyectos científicos.



## CONCLUSIONES

6





1. La B35DNAP posee capacidad de polimerización de DNA de manera procesiva y acoplada a desplazamiento de banda, sin necesidad de la presencia adicional de factores de procesividad o proteínas de tipo helicasas. Además, la B35DNAP tiene actividad exonucleasa 3'-5' con capacidad correctora de errores que actúa de manera coordinada con la actividad de polimerización. Esta actividad, junto a la alta especificidad de inserción del nucleótido complementario al molde mostrada por la B35DNAP, hacen que esta DNAP tenga una alta fidelidad de replicación, similar a la de otras DNAPs empleadas habitualmente para la amplificación del DNA.
2. El ORF4 del genoma de Bam35 codifica la TP utilizada como *primer* en la replicación de su TP-DNA. Así, la B35DNAP y la TP son necesarias y suficientes para la replicación completa del genoma de Bam35 *in vitro*.
3. El residuo conservado tirosina 194 de la TP de Bam35 proporciona el grupo OH necesario para la iniciación de la replicación. Por otro lado, la tirosina 172, también conservada en las secuencias relacionadas, tiene un papel en la interacción con la B35DNAP.
4. La iniciación de la replicación del genoma de Bam35 mediada por la TP está dirigida por el tercer nucleótido del extremo 3' de la cadena molde del DNA.
5. La información genética de los extremos 3' del genoma de Bam35 puede recuperarse mediante un mecanismo de *jumping-back* de un nucleótido.
6. La B35DNAP es capaz de replicar de manera procesiva un molde de DNA con sitios abásicos introduciendo, preferentemente, una adenina enfrente del daño. Tanto la inserción del nucleótido enfrente del daño como la posterior extensión del *primer* se ven influenciados por el contexto de secuencia del DNA.
7. Al igual que en otras DNAPs replicativas, la capacidad de síntesis de un molde de DNA con sitios abásicos por la B35DNAP se ve contrarrestada por su actividad exonucleasa 3'-5' hasta que el error en el apareamiento producido por el nucleótido insertado enfrente del sitio abásico se encuentra a 5 nucleótidos *upstream* del extremo 3' del *primer*. La eliminación del subdominio TPR2, requerido para la procesividad, impide la extensión del *primer* más allá del daño.

8. Con el sistema de doble híbrido de levaduras se detectaron interacciones entre 25 proteínas codificadas por el genoma de Bam35, de las cuales 12 eran de función desconocida. Los resultados nos permiten sugerir que, en el virión de Bam35, las proteínas P17 y P24 podrían ser la proteína minoritaria de la cápsida y la formadora del pentón, respectivamente. Por otra parte, la proteína de función desconocida P22 probablemente esté asociada al vértice especial.
9. Con respecto a las proteínas con dominios de transmembrana, nuestros datos sugieren que la proteína lítica P30 y la P32 estarían localizadas en la membrana lipídica a través de la interacción con la proteína de membrana P26. Las interacciones mostradas con la proteína P26 sugieren un papel estructural esencial. Por otro lado, la proteína de función desconocida P15 probablemente esté asociadas al vértice especial. Finalmente, la proteína P20 podría ser la responsable de la estabilización de las espículas.
10. La interacción de la TP de Bam35 consigo misma podría ser esencial para el reconocimiento de los orígenes de replicación del genoma de Bam35. Del mismo modo, la interacción de la TP con la proteína P1 sugieren un papel de esta proteína en la iniciación de la replicación, apoyado por su capacidad de unión a DNA.



## CONCLUSIONS



1. The B35DNAP is endowed with processive DNA polymerization coupled to strand displacement without the requirement of the additional presence of processivity factors or helicases-like proteins. Besides, the B35DNAP shows a 3'-5' exonuclease activity with a proofreading capacity that acts in a coordinated manner with the polymerization activity. This activity together with the high specificity of insertion of the nucleotide dictated by the template strand provides to the B35DNAP a high replication fidelity, similar to that of DNA polymerases generally used to amplify DNA.
2. The ORF4 of the Bam35 genome encodes a TP used as a primer in the TP-DNA replication. Thus, the B35DNAP and the TP are necessary and sufficient to accomplish the full-length Bam35 genome replication.
3. The conserved residue tyrosine 194 of the Bam35 TP provides the OH group necessary for the initiation of the replication. On the other hand, tyrosine 172, also conserved among Bam35 related TPs, is involved in the interaction with the B35DNAP.
4. The initiation of Bam35 genome replication TP-primed is directed by the third nucleotide at the 3' end of the DNA template strand.
5. The genetic information of the Bam35 genome 3' ends can be recovered through a single-nucleotide jumping-back mechanism.
6. The B35DNAP is able to replicate in a processive manner a DNA template with abasic sites through the insertion, preferably, of an adenine in front of the damage. Both, the nucleotide insertion in front of the damage and the latter primer extension are influenced by the DNA sequence context.
7. Similarly to other DNA polymerases, the capacity of B35DNAP to replicate a DNA template with abasic sites is counteracted by its 3'-5' exonuclease activity until the nucleotide:abasic site mismatch is located at 5 nucleotides upstream from the primer 3' end. The elimination of the TPR2 subdomain, required for processivity, impairs primer extension beyond the abasic site.

8. With the yeast two-hybrid screening we detected interactions between 25 proteins encoded by the Bam35 genome, 12 of which had unknown functions. The results suggest that, in the Bam35 virion, the P17 and P24 proteins could be the minor capsid protein and the penton protein, respectively. On the other hand, the P22 protein of unknown function is probably associated with the special vertex.
9. Regarding the proteins with transmembrane domains, our data suggest that the P30 lytic protein and the P32 would be located in the inner membrane of Bam35, probably, through the interaction with P26. The P26 interactions suggest an essential structural role. On the other hand, the P15 protein of unknown function is, probably, associated with the special vertex. Finally, the P20 protein could be responsible of the spikes stabilization.
10. The Bam35 TP interaction with itself could be essential for the recognition of the replication origin of the Bam35 genome. Similarly, the interaction between TP and P1 proteins suggest a role of P1 in the initiation of the replication, which is supported by its DNA binding capacity.



## **BIBLIOGRAFÍA**

7





- Abbotts, J., SenGupta, D. N., Zmudzka, B., Widen, S. G., Notario, V. y Wilson, S. H. 1988. Expression of human DNA polymerase  $\beta$  in *Escherichia coli* and characterization of the recombinant enzyme. *Biochemistry*, 27, 901-909.
- Abrescia, N. G., Cockburn, J. J., Grimes, J. M., Sutton, G. C., Diprose, J. M., Butcher, S. J., Fuller, S. D., San Martín, C., Burnett, R. M., Stuart, D. I., Bamford, D. H. y Bamford, J. K. 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature*, 432, 68-74.
- Abulreesh, H. H., Osman, G. E. y Assaeedi, A. S. 2012. Characterization of insecticidal genes of *Bacillus thuringiensis* strains isolated from arid environments. *Indian J Microbiol*, 52, 500-503.
- Andreev, Y. A., Kozlov, S. A., Vassilevski, A. A. y Grishin, E. V. 2010. Cyanogen bromide cleavage of proteins in salt and buffer solutions. *Anal Biochem*, 407, 144-146.
- Arana, M. E., Seki, M., Wood, R. D., Rogozin, I. B. y Kunkel, T. A. 2008. Low-fidelity DNA synthesis by human DNA polymerase  $\theta$ . *Nucleic Acids Res*, 36, 3847-3856.
- Ballesteros-Plaza, D., Holguera, I., Scheffers, D. J., Salas, M. y Muñoz-Espín, D. 2013. Phage  $\Phi$ 29 protein p1 promotes replication by associating with the FtsZ ring of the divisome in *Bacillus subtilis*. *Proc Natl Acad Sci U S A*, 110, 12313-12318.
- Bamford, D. H., Ravantti, J. J., Ronnholm, G., Laurinavicius, S., Kukkaro, P., Dyll-Smith, M., Somerharju, P., Kalkkinen, N. y Bamford, J. K. 2005. Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon *Haloarcula hispanica*. *J Virol*, 79, 9097-9107.
- Baños, B., Lázaro, J. M., Villar, L., Salas, M. y de Vega, M. 2008. Characterization of a *Bacillus subtilis* 64-kDa DNA polymerase X potentially involved in DNA repair. *J Mol Biol*, 384, 1019-1028.
- Bath, C., Cukalac, T., Porter, K. y Dyll-Smith, M. L. 2006. His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, *Salterprovirus*. *Virology*, 350, 228-239.
- Beckman, R. A., Mildvan, A. S. y Loeb, L. A. 1985. On the fidelity of DNA replication: manganese mutagenesis *in vitro*. *Biochemistry*, 24, 5810-5817.

Beese, L. S., Derbyshire, V. y Steitz, T. A. 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science*, 260, 352-355.

Benson, S. D., Bamford, J. K., Bamford, D. H. y Burnett, R. M. 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell*, 98, 825-833.

Berman, A. J., Kamtekar, S., Goodman, J. L., Lázaro, J. M., de Vega, M., Blanco, L., Salas, M. y Steitz, T. A. 2007. Structures of  $\Phi$ 29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J*, 26, 3494-3505.

Bernad, A., Blanco, L., Lázaro, J. M., Martín, G. y Salas, M. 1989. A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell*, 59, 219-228.

Bernad, A., Lázaro, J. M., Salas, M. y Blanco, L. 1990. The highly conserved amino acid sequence motif Tyr-Gly-Asp-Thr-Asp-Ser in  $\alpha$ -like DNA polymerases is required by phage  $\Phi$ 29 DNA polymerase for protein-primed initiation and polymerization. *Proc Natl Acad Sci U S A*, 87, 4610-4614.

Blanco, L., Gutiérrez, J., Lázaro, J. M., Bernad, A. y Salas, M. 1986. Replication of phage  $\Phi$ 29 DNA *in vitro*: role of the viral protein p6 in initiation and elongation. *Nucleic Acids Res*, 14, 4923-4937.

Blanco, L., Bernad, A., Lázaro, J. M., Martín, G., Garmendia, C. y Salas, M. 1989. Highly efficient DNA synthesis by the phage  $\Phi$ 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem*, 264, 8935-8940.

Blanco, L., Bernad, A., Blasco, M. A. y Salas, M. 1991. A general structure for DNA-dependent DNA polymerases. *Gene*, 100, 27-38.

Blanco, L., Bernad, A., Esteban, J. A. y Salas, M. 1992. DNA-independent deoxynucleotidylation of the  $\Phi$ 29 terminal protein by the  $\Phi$ 29 DNA polymerase. *J Biol Chem*, 267, 1225-1230.

Blasche, S., Wuchty, S., Rajagopala, S. V. y Uetz, P. 2013. The protein interaction network of bacteriophage  $\lambda$  with its host, *Escherichia coli*. *J Virol*, 87, 12745-12755.

- Blasco, M. A., Blanco, L., Pares, E., Salas, M. y Bernad, A. 1990. Structural and functional analysis of temperature-sensitive mutants of the phage  $\Phi$ 29 DNA polymerase. *Nucleic Acids Res*, 18, 4763-4770.
- Blasco, M. A., Lázaro, J. M., Blanco, L. y Salas, M. 1993.  $\Phi$ 29 DNA polymerase active site. The conserved amino acid motif "Kx<sub>3</sub>NSxYG" is involved in template-primer binding and dNTP selection. *J Biol Chem*, 268, 16763-16770.
- Blasco, M. A., Méndez, J., Lázaro, J. M., Blanco, L. y Salas, M. 1995. Primer terminus stabilization at the  $\Phi$ 29 DNA polymerase active site. Mutational analysis of conserved motif KxY. *J Biol Chem*, 270, 2735-2740.
- Bogenhagen, D. F., Pinz, K. G. y Pérez-Jannotti, R. M. 2001. Enzymology of mitochondrial base excision repair. *Prog Nucleic Acid Res Mol Biol*, 68, 257-271.
- Bollum, F. J. 1960. Oligodeoxyribonucleotide primers for calf thymus polymerase. *J Biol Chem*, 235, 18-20.
- Boule, J. B., Rougeon, F. y Papanicolaou, C. 2001. Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides. *J Biol Chem*, 276, 31388-31393.
- Braithwaite, D. K. y Ito, J. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res*, 21, 787-802.
- Bruckner, A., Polge, C., Lentze, N., Auerbach, D. y Schlattner, U. 2009. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*, 10, 2763-2788.
- Caldentey, J., Blanco, L., Savilahti, H., Bamford, D. H. y Salas, M. 1992. *In vitro* replication of bacteriophage PRD1 DNA. Metal activation of protein-primed initiation and DNA elongation. *Nucleic Acids Res*, 20, 3971-3976.
- Caldentey, J., Blanco, L., Bamford, D. H. y Salas, M. 1993. *In vitro* replication of bacteriophage PRD1 DNA. Characterization of the protein-primed initiation site. *Nucleic Acids Res*, 21, 3725-3730.
- Caldentey, J., Hanninen, A. L. y Bamford, D. H. 1994. Gene XV of bacteriophage PRD1 encodes a lytic enzyme with muramidase activity. *Eur J Biochem*, 225, 341-346.

Capson, T. L., Peliska, J. A., Kaboord, B. F., Frey, M. W., Lively, C., Dahlberg, M. y Benkovic, S. J. 1992. Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. *Biochemistry*, 31, 10984-10994.

Carthew, R. W., Chodosh, L. A. y Sharp, P. A. 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell*, 43, 439-448.

Cepko, C., Ryder, E., Fekete, D. M. y Bruhn, S. 1998. Detection of  $\beta$ -galactosidase and alkaline phosphatase activities in tissue. In: Spector D.L., G. R. D. y L. L. A. (ed.) *Cells: A Laboratory Manual, Volume 3: Subcellular Location of Genes and Their Products*. Cold Spring Harbor Cold Spring Harbor Laboratory Press.

Chan, S. H., Yu, A. M. y McVey, M. 2010. Dual roles for DNA polymerase  $\theta$  in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet*, 6, e1001005.

Choi, J. Y., Lim, S., Kim, E. J., Jo, A. y Guengerich, F. P. 2010. Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases  $\alpha$ ,  $\Delta$ ,  $\eta$ ,  $\iota$ ,  $\kappa$ , and REV1. *J Mol Biol*, 404, 34-44.

Cline, J., Braman, J. C. y Hogrefe, H. H. 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res*, 24, 3546-3551.

Cyrklaff, M., Risco, C., Fernández, J. J., Jiménez, M. V., Esteban, M., Baumeister, W. y Carrascosa, J. L. 2005. Cryo-electron tomography of vaccinia virus. *Proc Natl Acad Sci U S A*, 102, 2772-2777.

Daugelavicius, R., Gaidelyte, A., Cvirkaitė-Krupovic, V. y Bamford, D. H. 2007. On-line monitoring of changes in host cell physiology during the one-step growth cycle of *Bacillus* phage Bam35. *J Microbiol Methods*, 69, 174-179.

de Jong, R. N., Meijer, L. A. y van der Vliet, P. C. 2003. DNA binding properties of the adenovirus DNA replication priming protein pTP. *Nucleic Acids Res*, 31, 3274-3286.

De Vega, M. y Salas, M. 2007. A highly conserved Tyrosine residue of family B DNA polymerases contributes to dictate translesion synthesis past 8-oxo-7,8-dihydro-2'-deoxyguanosine. *Nucleic Acids Res*, 35, 5096-5107.

- De Vega, M. y Salas, M. 2011. Protein-primed replication of bacteriophage  $\Phi$ 29 DNA. In: Kusic-Tisma, J. (ed.) *DNA replication and related cellular processes*. Intech Croacia.
- Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M. y Lasken, R. S. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A*, 99, 5261-5266.
- Dekker, J., Kanellopoulos, P. N., Loonstra, A. K., van Oosterhout, J. A., Leonard, K., Tucker, P. A. y van der Vliet, P. C. 1997. Multimerization of the adenovirus DNA-binding protein is the driving force for ATP-independent DNA unwinding during strand displacement synthesis. *EMBO J*, 16, 1455-1463.
- Del Prado, A., Villar, L., de Vega, M. y Salas, M. 2012. Involvement of residues of the  $\Phi$ 29 terminal protein intermediate and priming domains in the formation of a stable and functional heterodimer with the replicative DNA polymerase. *Nucleic Acids Res*, 40, 3886-3897.
- Del Prado, A., Lázaro, J. M., Longás, E., Villar, L., de Vega, M. y Salas, M. 2015. Insights into the determination of the templating nucleotide at the initiation of  $\Phi$ 29 DNA replication. *J Biol Chem*, 290, 27138-27145.
- Delattre, H., Souiai, O., Fagoonee, K., Guerois, R. y Petit, M. A. 2016. Phagonaute: A web-based interface for phage synteny browsing and protein function prediction. *Virology*, 496, 42-50.
- Desnues, C., Boyer, M. y Raoult, D. 2012. Sputnik, a virophage infecting the viral domain of life. *Adv Virus Res*, 82, 63-89.
- Domínguez, O., Ruiz, J. F., Lain de Lera, T., García-Díaz, M., González, M. A., Kirchhoff, T., Martínez, A. C., Bernad, A. y Blanco, L. 2000. DNA polymerase  $\mu$  (Pol  $\mu$ ), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J*, 19, 1731-1742.
- Dufour, E., Méndez, J., Lázaro, J. M., de Vega, M., Blanco, L. y Salas, M. 2000. An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. *J Mol Biol*, 304, 289-300.

- Esteban, J. A., Bernad, A., Salas, M. y Blanco, L. 1992. Metal activation of synthetic and degradative activities of  $\Phi$ 29 DNA polymerase, a model enzyme for protein-primed DNA replication. *Biochemistry*, 31, 350-359.
- Esteban, J. A., Salas, M. y Blanco, L. 1993. Fidelity of  $\Phi$ 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J Biol Chem*, 268, 2719-2726.
- Fields, S. y Song, O. 1989. A novel genetic system to detect protein-protein interactions. *Nature*, 340, 245-246.
- Fischer, M. G. y Suttle, C. A. 2011. A virophage at the origin of large DNA transposons. *Science*, 332, 231-234.
- Fischetti, V. A. 2011. Exploiting what phage have evolved to control gram-positive pathogens. *Bacteriophage*, 1, 188-194.
- Fornelos, N., Bamford, J. K. y Mahillon, J. 2011. Phage-borne factors and host LexA regulate the lytic switch in phage GIL01. *J Bacteriol*, 193, 6008-6019.
- Fornelos, N., Butala, M., Hodnik, V., Anderluh, G., Bamford, J. K. y Salas, M. 2015. Bacteriophage GIL01 gp7 interacts with host LexA repressor to enhance DNA binding and inhibit RecA-mediated auto-cleavage. *Nucleic Acids Res*, 43, 7315-7329.
- Forterre, P. y Prangishvili, D. 2009. The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann N Y Acad Sci*, 1178, 65-77.
- Fortune, J. M., Pavlov, Y. I., Welch, C. M., Johansson, E., Burgers, P. M. y Kunkel, T. A. 2005. *Saccharomyces cerevisiae* DNA polymerase  $\Delta$ : high fidelity for base substitutions but lower fidelity for single- and multi-base deletions. *J Biol Chem*, 280, 29980-29987.
- Freudenthal, B. D., Beard, W. A. y Wilson, S. H. 2012. Structures of dNTP intermediate states during DNA polymerase active site assembly. *Structure*, 20, 1829-1837.

- Gaidelyte, A., Jaatinen, S. T., Daugelavicius, R., Bamford, J. K. y Bamford, D. H. 2005. The linear double-stranded DNA of phage Bam35 enters lysogenic host cells, but the late phage functions are suppressed. *J Bacteriol*, 187, 3521-3527.
- Gaidelyte, A., Cvirkaitė-Krupovic, V., Daugelavicius, R., Bamford, J. K. y Bamford, D. H. 2006. The entry mechanism of membrane-containing phage Bam35 infecting *Bacillus thuringiensis*. *J Bacteriol*, 188, 5925-5934.
- Ganai, R. A. y Johansson, E. 2016. DNA replication-A matter of fidelity. *Mol Cell*, 62, 745-755.
- García-Díaz, M., Domínguez, O., López-Fernández, L. A., de Lera, L. T., Saniger, M. L., Ruiz, J. F., Párraga, M., García-Ortiz, M. J., Kirchhoff, T., del Mazo, J., Bernad, A. y Blanco, L. 2000. DNA polymerase  $\lambda$  (Pol  $\lambda$ ), a novel eukaryotic DNA polymerase with a potential role in meiosis. *J Mol Biol*, 301, 851-867.
- García-Díaz, M. y Kunkel, T. A. 2006. Mechanism of a genetic glissando: structural biology of indel mutations. *Trends Biochem Sci*, 31, 206-214.
- García-Gómez, S., Reyes, A., Martínez-Jiménez, M. I., Chocrón, E. S., Mouron, S., Terrados, G., Powell, C., Salido, E., Méndez, J., Holt, I. J. y Blanco, L. 2013. PrimPol, an archaic primase/polymerase operating in human cells. *Mol Cell*, 52, 541-553.
- Gella, P., Salas, M. y Mencía, M. 2014. Improved artificial origins for phage  $\Phi$ 29 terminal protein-primed replication. Insights into early replication events. *Nucleic Acids Res*, 42, 9792-9806.
- Gillis, A. y Mahillon, J. 2014a. Prevalence, genetic diversity, and host range of tectiviruses among members of the *Bacillus cereus* group. *Appl Environ Microbiol*, 80, 4138-4152.
- Gillis, A. y Mahillon, J. 2014b. Phages preying on *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*: past, present and future. *Viruses*, 6, 2623-2672.
- González-Huici, V., Lázaro, J. M., Salas, M. y Hermoso, J. M. 2000. Specific recognition of parental terminal protein by DNA polymerase for initiation of protein-primed DNA replication. *J Biol Chem*, 275, 14678-14683.

Goodier, J. L. y Kazazian, H. H., Jr. 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*, 135, 23-35.

Goodman, M. F. y Woodgate, R. 2013. Translesion DNA polymerases. *Cold Spring Harb Perspect Biol*, 5, a010363.

Gotte, M., Li, X. y Wainberg, M. A. 1999. HIV-1 reverse transcription: a brief overview focused on structure-function relationships among molecules involved in initiation of the reaction. *Arch Biochem Biophys*, 365, 199-210.

Grahn, A. M., Daugelavicius, R. y Bamford, D. H. 2002. Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. *Mol Microbiol*, 46, 1199-1209.

Graziewicz, M. A., Longley, M. J. y Copeland, W. C. 2006. DNA polymerase  $\gamma$  in mitochondrial DNA replication and repair. *Chem Rev*, 106, 383-405.

Grazziotin, A. L., Koonin, E. V. y Kristensen, D. M. 2016. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res*, pii: gkw975.

Hauser, R., Sabri, M., Moineau, S. y Uetz, P. 2011. The proteome and interactome of *Streptococcus pneumoniae* phage Cp-1. *J Bacteriol*, 193, 3135-3138.

Hauser, R., Blasche, S., Dokland, T., Haggard-Ljungquist, E., von Brunn, A., Salas, M., Casjens, S., Molineux, I. y Uetz, P. 2012. Bacteriophage protein-protein interactions. *Adv Virus Res*, 83, 219-298.

Hermoso, J. M., Méndez, E., Soriano, F. y Salas, M. 1985. Location of the serine residue involved in the linkage between the terminal protein and the DNA of phage  $\Phi$ 29. *Nucleic Acids Res*, 13, 7715-7728.

Hoeben, R. C. y Uil, T. G. 2013. Adenovirus DNA replication. *Cold Spring Harb Perspect Biol*, 5, a013003.

Hogg, M., Wallace, S. S. y Doublié, S. 2004. Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site. *EMBO J*, 23, 1483-1493.



- Hogg, M., Seki, M., Wood, R. D., Doublié, S. y Wallace, S. S. 2011. Lesion bypass activity of DNA polymerase  $\theta$  (POLQ) is an intrinsic property of the pol domain and depends on unique sequence inserts. *J Mol Biol*, 405, 642-652.
- Holkers, M., Maggio, I., Henriques, S. F., Janssen, J. M., Cathomen, T. y Goncalves, M. A. 2014. Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nat Methods*, 11, 1051-1057.
- Illana, B., Blanco, L. y Salas, M. 1996. Functional characterization of the genes coding for the terminal protein and DNA polymerase from bacteriophage GA-1. Evidence for a sliding-back mechanism during protein-primed GA-1 DNA replication. *J Mol Biol*, 264, 453-464.
- Illana, B., Zaballos, A., Blanco, L. y Salas, M. 1998. The RGD sequence in phage  $\Phi$ 29 terminal protein is required for interaction with  $\Phi$ 29 DNA polymerase. *Virology*, 248, 12-19.
- Ito, J. y Braithwaite, D. K. 1991. Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res*, 19, 4045-4057.
- Iyer, L. M., Koonin, E. V., Leipe, D. D. y Aravind, L. 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res*, 33, 3875-3896.
- Jaatinen, S. T., Viitanen, S. J., Bamford, D. H. y Bamford, J. K. 2004. Integral membrane protein P16 of bacteriophage PRD1 stabilizes the adsorption vertex structure. *J Virol*, 78, 9790-9797.
- Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. y Arnold, E. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc Natl Acad Sci U S A*, 90, 6320-6324.
- Jalasvuori, M., Palmu, S., Gillis, A., Kokko, H., Mahillon, J., Bamford, J. K. y Fornelos, N. 2013. Identification of five novel tectiviruses in *Bacillus* strains: analysis of a highly variable region generating genetic diversity. *Res Microbiol*, 164, 118-126.

- Johansson, E. y Dixon, N. 2013. Replicative DNA polymerases. *Cold Spring Harb Perspect Biol*, 5, a012799.
- Johnson, S. J. y Beese, L. S. 2004. Structures of mismatch replication errors observed in a DNA polymerase. *Cell*, 116, 803-816.
- Joyce, C. M. y Steitz, T. A. 1994. Function and structure relationships in DNA polymerases. *Annu Rev Biochem*, 63, 777-822.
- Joyce, C. M. y Steitz, T. A. 1995. Polymerase structures and function: variations on a theme? *J Bacteriol*, 177, 6321-6329.
- Jozwiakowski, S. K. y Connolly, B. A. 2009. Plasmid-based lacZ $\alpha$  assay for DNA polymerase fidelity: application to archaeal family-B DNA polymerase. *Nucleic Acids Res*, 37, e102.
- Jozwiakowski, S. K., Keith, B. J., Gilroy, L., Doherty, A. J. y Connolly, B. A. 2014. An archaeal family-B DNA polymerase variant able to replicate past DNA damage: occurrence of replicative and translesion synthesis polymerases within the B family. *Nucleic Acids Res*, 42, 9949-9963.
- Kamtekar, S., Berman, A. J., Wang, J., Lázaro, J. M., de Vega, M., Blanco, L., Salas, M. y Steitz, T. A. 2004. Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage  $\Phi$ 29. *Mol Cell*, 16, 609-618.
- Kan, S., Fornelos, N., Schuch, R. y Fischetti, V. A. 2013. Identification of a ligand on the Wip1 bacteriophage highly specific for a receptor on *Bacillus anthracis*. *J Bacteriol*, 195, 4355-4364.
- Kapitonov, V. V. y Jurka, J. 2006. Self-synthesizing DNA transposons in eukaryotes. *Proc Natl Acad Sci U S A*, 103, 4540-4545.
- Karhu, N. J., Ziedaite, G., Bamford, D. H. y Bamford, J. K. 2007. Efficient DNA packaging of bacteriophage PRD1 requires the unique vertex protein P6. *J Virol*, 81, 2970-2979.
- Karttunen, J., Mantynen, S., Ihalainen, T. O., Bamford, J. K. y Oksanen, H. M. 2015. Non-structural proteins P17 and P33 are involved in the assembly of the internal membrane-containing virus PRD1. *Virology*, 482, 225-233.

- Kazazian, H. H., Jr. 2004. Mobile elements: drivers of genome evolution. *Science*, 303, 1626-1632.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. y Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647-1649.
- Kent, T., Mateos-Gómez, P. A., Sfeir, A. y Pomerantz, R. T. 2016. Polymerase  $\theta$  is a robust terminal transferase that oscillates between three different mechanisms during end-joining. *Elife*, 5, e13740.
- King, A. J. y van der Vliet, P. C. 1994. A precursor terminal protein-trinucleotide intermediate during initiation of adenovirus DNA replication: regeneration of molecular ends *in vitro* by a jumping back mechanism. *EMBO J*, 13, 5786-5792.
- Klose, T. y Rossmann, M. G. 2014. Structure of large dsDNA viruses. *Biol Chem*, 395, 711-719.
- Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. y Steitz, T. A. 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, 256, 1783-1790.
- Kokoska, R. J., McCulloch, S. D. y Kunkel, T. A. 2003. The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase  $\eta$  and *Sulfolobus solfataricus* Dpo4. *J Biol Chem*, 278, 50537-50545.
- Koonin, E. V. y Dolja, V. V. 2013. A virocentric perspective on the evolution of life. *Curr Opin Virol*, 3, 546-557.
- Koonin, E. V. y Dolja, V. V. 2014. Virus world as an evolutionary network of *viruses* and capsidless selfish elements. *Microbiol Mol Biol Rev*, 78, 278-303.
- Krogh, A., Larsson, B., von Heijne, G. y Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, 305, 567-580.
- Krupovic, M. y Bamford, D. H. 2010. Order to the viral universe. *J Virol*, 84, 12476-12479.

Krupovic, M. y Bamford, D. H. 2011. Double-stranded DNA *viruses*: 20 families and only five different architectural principles for virion assembly. *Curr Opin Virol*, 1, 118-124.

Krupovic, M., Prangishvili, D., Hendrix, R. W. y Bamford, D. H. 2011. Genomics of bacterial and archaeal *viruses*: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev*, 75, 610-635.

Krupovic, M. y Koonin, E. V. 2015. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat Rev Microbiol*, 13, 105-115.

Kunkel, T. A., Roberts, J. D. y Sugino, A. 1991. The fidelity of DNA synthesis by the catalytic subunit of yeast DNA polymerase  $\alpha$  alone and with accessory proteins. *Mutat Res*, 250, 175-182.

Kunkel, T. A. 2004. DNA replication fidelity. *J Biol Chem*, 279, 16895-16898.

Kunkel, T. A. y Erie, D. A. 2015. Eukaryotic mismatch repair in relation to DNA replication. *Annu Rev Genet*, 49, 291-313.

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E. y Raoult, D. 2008. The virophage as a unique parasite of the giant mimivirus. *Nature*, 455, 100-104.

Laurinavicius, S., Bamford, D. H. y Somerharju, P. 2007. Transbilayer distribution of phospholipids in bacteriophage membranes. *Biochim Biophys Acta*, 1768, 2568-2577.

Laurinmaki, P. A., Huiskonen, J. T., Bamford, D. H. y Butcher, S. J. 2005. Membrane proteins modulate the bilayer curvature in the bacterial virus Bam35. *Structure*, 13, 1819-1828.

Lázaro, J. M., Blanco, L. y Salas, M. 1995. Purification of bacteriophage  $\Phi$ 29 DNA polymerase. *Methods Enzymol*, 262, 42-49.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature*, 362, 709-715.

- Liu, B., Xue, Q., Tang, Y., Cao, J., Guengerich, F. P. y Zhang, H. 2016. Mechanisms of mutagenesis: DNA replication in the presence of DNA damage. *Mutat Res-Rev Mutat Res*, 768, 53-67.
- Longás, E., de Vega, M., Lázaro, J. M. y Salas, M. 2006. Functional characterization of highly processive protein-primed DNA polymerases from phages Nf and GA-1, endowed with a potent strand displacement capacity. *Nucleic Acids Res*, 34, 6051-6063.
- Maga, G., Shevelev, I., Ramadan, K., Spadari, S. y Hubscher, U. 2002. DNA polymerase  $\theta$  purified from human cells is a high-fidelity enzyme. *J Mol Biol*, 319, 359-369.
- Maor-Shoshani, A., Hayashi, K., Ohmori, H. y Livneh, Z. 2003. Analysis of translesion replication across an abasic site by DNA polymerase IV of *Escherichia coli*. *DNA Repair (Amst)*, 2, 1227-1238.
- Mariano, R., Wuchty, S., Vizoso-Pinto, M. G., Hauser, R. y Uetz, P. 2016. The interactome of *Streptococcus pneumoniae* and its bacteriophages show highly specific patterns of interactions among bacteria and their phages. *Sci Rep*, 6, 24597.
- Martín, G., Lázaro, J. M., Méndez, E. y Salas, M. 1989. Characterization of the phage  $\Phi$ 29 protein p5 as a single-stranded DNA binding protein. Function in  $\Phi$ 29 DNA-protein p3 replication. *Nucleic Acids Res*, 17, 3663-3672.
- Martínez-Jiménez, M. I., García-Gómez, S., Bebenek, K., Sastre-Moreno, G., Calvo, P. A., Díaz-Talavera, A., Kunkel, T. A. y Blanco, L. 2015. Alternative solutions and new scenarios for translesion DNA synthesis by human PrimPol. *DNA Repair (Amst)*, 29, 127-138.
- Mattila, S., Oksanen, H. M. y Bamford, J. K. 2015. Probing protein interactions in the membrane-containing virus PRD1. *J Gen Virol*, 96, 453-462.
- McCulloch, S. D. y Kunkel, T. A. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*, 18, 148-161.
- McCulloch, S. D., Kokoska, R. J., Garg, P., Burgers, P. M. y Kunkel, T. A. 2009. The efficiency and fidelity of 8-oxo-guanine bypass by DNA polymerases  $\Delta$  and  $\eta$ . *Nucleic Acids Res*, 37, 2830-2840.

McDonell, M. W., Simon, M. N. y Studier, F. W. 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J Mol Biol*, 110, 119-146.

McInerney, P., Adams, P. y Hadi, M. Z. 2014. Error rate comparison during polymerase chain reaction by DNA polymerase. *Mol Biol Int*, 2014, 287430.

Mehla, J., Dedrick, R. M., Caufield, J. H., Sieftring, R., Mair, M., Johnson, A., Hatfull, G. F. y Uetz, P. 2015a. The protein interactome of mycobacteriophage Giles predicts functions for unknown proteins. *J Bacteriol*, 197, 2508-2516.

Mehla, J., Caufield, J. H. y Uetz, P. 2015b. Mapping protein-protein interactions using yeast two-hybrid assays. *Cold Spring Harb Protoc*, 2015, 442-452.

Meijer, W. J., Horcajadas, J. A. y Salas, M. 2001.  $\Phi$ 29 family of phages. *Microbiol Mol Biol Rev*, 65, 261-287.

Mencía, M., Gella, P., Camacho, A., de Vega, M. y Salas, M. 2011. Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage  $\Phi$ 29. *Proc Natl Acad Sci U S A*, 108, 18655-18660.

Méndez, J., Blanco, L., Esteban, J. A., Bernad, A. y Salas, M. 1992. Initiation of  $\Phi$ 29 DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. *Proc Natl Acad Sci U S A*, 89, 9579-9583.

Miller, H. y Grollman, A. P. 1997. Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry*, 36, 15336-15342.

Mindich, L., Bamford, D., Goldthwaite, C., Laverty, M. y Mackenzie, G. 1982. Isolation of nonsense mutants of lipid-containing bacteriophage PRD1. *J Virol*, 44, 1013-1020.

Minnick, D. T., Astatke, M., Joyce, C. M. y Kunkel, T. A. 1996. A thumb subdomain mutant of the large fragment of *Escherichia coli* DNA polymerase I with reduced DNA binding affinity, processivity, and frameshift fidelity. *J Biol Chem*, 271, 24954-24961.

- Moelling, K., Bolognesi, D. P. y Bauer, H. 1971. Polypeptides of avian RNA tumor viruses. 3. Purification and identification of a DNA synthesizing enzyme. *Virology*, 45, 298-302.
- Moore, C. M. y Helmann, J. D. 2005. Metal ion homeostasis in *Bacillus subtilis*. *Curr Opin Microbiol*, 8, 188-195.
- Morhoshi, F. y Munakata, N. 1995. Diverse capacities for the adaptive response to DNA alkylation in *Bacillus* species and strains. *Mutat Res*, 337, 97-110.
- Nagy, E. 1974. A highly specific phage attacking *Bacillus anthracis* strain Sterne. *Acta Microbiol Acad Sci Hung*, 21, 257-263.
- Nakamura, J. y Swenberg, J. A. 1999. Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer Res*, 59, 2522-2526.
- Nelson, J. R., Cai, Y. C., Giesler, T. L., Farchaus, J. W., Sundaram, S. T., Ortiz-Rivera, M., Hosta, L. P., Hewitt, P. L., Mamone, J. A., Palaniappan, C. y Fuller, C. W. 2002. TempliPhi,  $\Phi$ 29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *Biotechniques*, Suppl, 44-47.
- Nelson, J. R. 2014. Random-primed,  $\Phi$ 29 DNA polymerase-based whole genome amplification. *Curr Protoc Mol Biol*, 105, Unit 15 13.
- Ng, L., Weiss, S. J. y Fisher, P. A. 1989. Recognition and binding of template-primers containing defined abasic sites by *Drosophila* DNA polymerase  $\alpha$  holoenzyme. *J Biol Chem*, 264, 13018-13023.
- Niimi, A., Limsirichaikul, S., Yoshida, S., Iwai, S., Masutani, C., Hanaoka, F., Kool, E. T., Nishiyama, Y. y Suzuki, M. 2004. Palm mutants in DNA polymerases  $\alpha$  and  $\eta$  alter DNA replication fidelity and translesion activity. *Mol Cell Biol*, 24, 2734-2746.
- O'Donnell, M., Langston, L. y Stillman, B. 2013. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol*, 5, a010108.

- Ohashi, E., Ogi, T., Kusumoto, R., Iwai, S., Masutani, C., Hanaoka, F. y Ohmori, H. 2000. Error-prone bypass of certain DNA lesions by the human DNA polymerase  $\kappa$ . *Genes Dev*, 14, 1589-1594.
- Oksanen, H. M. y Bamford, D. H. 2012. Family *Tectiviridae*. In: Andrew M. Q. King, M. J. A., Eric B. Carstens y Elliot J. Lefkowitz (ed.) *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego, CA.
- Oliveros, M., Yanez, R. J., Salas, M. L., Salas, J., Viñuela, E. y Blanco, L. 1997. Characterization of an African swine fever virus 20-kDa DNA polymerase involved in DNA repair. *J Biol Chem*, 272, 30899-30910.
- Olsen, R. H., Siak, J. S. y Gray, R. H. 1974. Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. *J Virol*, 14, 689-699.
- Ortín, J., Viñuela, E., Salas, M. y Vázquez, C. 1971. DNA-protein complex in circular DNA from phage  $\Phi$ 29. *Nat New Biol*, 234, 275-277.
- Osterman, A., Stellberger, T., Gebhardt, A., Kurz, M., Friedel, C. C., Uetz, P., Nitschko, H., Baiker, A. y Vizoso-Pinto, M. G. 2015. The Hepatitis E virus intraviral interactome. *Sci Rep*, 5, 13872.
- Pakula, T. M., Savilahti, H. y Bamford, D. H. 1989. Comparison of the amino acid sequence of the lytic enzyme from broad-host-range bacteriophage PRD1 with sequences of other *cell*-wall-peptidoglycan lytic enzymes. *Eur J Biochem*, 180, 149-152.
- Pakula, T. M., Caldentey, J., Serrano, M., Gutiérrez, C., Hermoso, J. M., Salas, M. y Bamford, D. H. 1990. Characterization of a DNA binding protein of bacteriophage PRD1 involved in DNA replication. *Nucleic Acids Res*, 18, 6553-6557.
- Pakula, T. M., Caldentey, J., Gutiérrez, C., Olkkonen, V. M., Salas, M. y Bamford, D. H. 1993. Overproduction, purification, and characterization of DNA-binding protein P19 of bacteriophage PRD1. *Gene*, 126, 99-104.
- Pastor-Palacios, G., López-Ramírez, V., Cardona-Félix, C. S. y Brieba, L. G. 2012. A transposon-derived DNA polymerase from *Entamoeba histolytica* displays intrinsic strand displacement, processivity and lesion bypass. *PLoS One*, 7, e49964.



Peng, X., Basta, T., Haring, M., Garrett, R. A. y Prangishvili, D. 2007. Genome of the Acidianus bottle-shaped virus and insights into the replication and packaging mechanisms. *Virology*, 364, 237-243.

Pérez-Arnaiz, P., Lázaro, J. M., Salas, M. y de Vega, M. 2006. Involvement of  $\Phi$ 29 DNA polymerase thumb subdomain in the proper coordination of synthesis and degradation during DNA replication. *Nucleic Acids Res*, 34, 3107-3115.

Pérez-Arnaiz, P., Longás, E., Villar, L., Lázaro, J. M., Salas, M. y de Vega, M. 2007. Involvement of phage  $\Phi$ 29 DNA polymerase and terminal protein subdomains in conferring specificity during initiation of protein-primed DNA replication. *Nucleic Acids Res*, 35, 7061-7073.

Quemin, E. R., Chlanda, P., Sachse, M., Forterre, P., Prangishvili, D. y Krupovic, M. 2016. Eukaryotic-like virus budding in *Archaea*. *MBio*, 7, e01439-01416.

Rajagopala, S. V., Casjens, S. y Uetz, P. 2011. The protein interaction map of bacteriophage  $\lambda$ . *BMC Microbiol*, 11, 213.

Rasband, W. S. 1997-2016. *National Institutes of Health, Bethesda, Maryland*.

Ravantti, J. J., Gaidelyte, A., Bamford, D. H. y Bamford, J. K. 2003. Comparative analysis of bacterial viruses Bam35, infecting a gram-positive host, and PRD1, infecting gram-negative hosts, demonstrates a viral lineage. *Virology*, 313, 401-414.

Redrejo-Rodríguez, M., Muñoz-Espín, D., Holguera, I., Mencía, M. y Salas, M. 2012. Functional eukaryotic nuclear localization signals are widespread in terminal proteins of bacteriophages. *Proc Natl Acad Sci U S A*, 109, 18482-18487.

Redrejo-Rodríguez, M., Muñoz-Espín, D., Holguera, I., Mencía, M. y Salas, M. 2013a. Nuclear and nucleoid localization are independently conserved functions in bacteriophage terminal proteins. *Mol Microbiol*, 90, 858-868.

Redrejo-Rodríguez, M., Muñoz-Espín, D., Holguera, I., Mencía, M. y Salas, M. 2013b. Nuclear localization signals in phage terminal proteins provide a novel gene delivery tool in mammalian cells. *Commun Integr Biol*, 6, e22829.

Rodríguez, I., Lázaro, J. M., Blanco, L., Kamtekar, S., Berman, A. J., Wang, J., Steitz, T. A., Salas, M. y de Vega, M. 2005. A specific subdomain in  $\Phi 29$  DNA polymerase confers both processivity and strand-displacement capacity. *Proc Natl Acad Sci U S A*, 102, 6407-6412.

Rodríguez, I., Lázaro, J. M., Salas, M. y de Vega, M. 2009. Involvement of the TPR2 subdomain movement in the activities of  $\Phi 29$  DNA polymerase. *Nucleic Acids Res*, 37, 193-203.

Rohwer, F. y Thurber, R. V. 2009. Viruses manipulate the marine environment. *Nature*, 459, 207-212.

Rothberg, P. G., Harris, T. J., Nomoto, A. y Wimmer, E. 1978. O<sup>4</sup>-(5'-uridylyl)tyrosine is the bond between the genome-linked protein and the RNA of poliovirus. *Proc Natl Acad Sci U S A*, 75, 4868-4872.

Ruiz, J. F., Juárez, R., García-Díaz, M., Terrados, G., Picher, A. J., González-Barrera, S., Fernández de Henestrosa, A. R. y Blanco, L. 2003. Lack of sugar discrimination by human Pol  $\mu$  requires a single glycine residue. *Nucleic Acids Res*, 31, 4441-4449.

Rydman, P. S., Caldentey, J., Butcher, S. J., Fuller, S. D., Rutten, T. y Bamford, D. H. 1999. Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *J Mol Biol*, 291, 575-587.

Rydman, P. S. y Bamford, D. H. 2000. Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. *Mol Microbiol*, 37, 356-363.

Rydman, P. S., Bamford, J. K. y Bamford, D. H. 2001. A minor capsid protein P30 is essential for bacteriophage PRD1 capsid assembly. *J Mol Biol*, 313, 785-795.

Sabouri, N. y Johansson, E. 2009. Translesion synthesis of abasic sites by yeast DNA polymerase  $\epsilon$ . *J Biol Chem*, 284, 31555-31563.

Salas, M., Mellado, R. P. y Viñuela, E. 1978. Characterization of a protein covalently linked to the 5' termini of the DNA of *Bacillus subtilis* phage  $\Phi 29$ . *J Mol Biol*, 119, 269-291.

Salas, M. 1991. Protein-priming of DNA replication. *Annu Rev Biochem*, 60, 39-71.

- Salas, M. 1999. Mechanisms of initiation of linear DNA replication in prokaryotes. *Genet Eng* (N Y), 21, 159-171.
- Salas, M. y de Vega, M. 2016. Protein-primed replication of bacteriophage  $\Phi$ 29 DNA. *Enzymes*, 39, 137-167.
- Salas, M. L. y Andrés, G. 2013. African swine fever virus morphogenesis. *Virus Res*, 173, 29-41.
- Sambrook, J. y Russell, D. 2001. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, NY.
- Saren, A. M., Ravantti, J. J., Benson, S. D., Burnett, R. M., Paulin, L., Bamford, D. H. y Bamford, J. K. 2005. A snapshot of viral evolution from genome analysis of the *Tectiviridae* family. *J Mol Biol*, 350, 427-440.
- Savilahti, H., Caldentey, J., Lundstrom, K., Syvaoja, J. E. y Bamford, D. H. 1991. Overexpression, purification, and characterization of *Escherichia coli* bacteriophage PRD1 DNA polymerase. *In vitro* synthesis of full-length PRD1 DNA with purified proteins. *J Biol Chem*, 266, 18737-18744.
- Schuch, R., Pelzek, A. J., Kan, S. y Fischetti, V. A. 2010. Prevalence of *Bacillus anthracis*-like organisms and bacteriophages in the intestinal tract of the earthworm *Eisenia fetida*. *Appl Environ Microbiol*, 76, 2286-2294.
- Serna-Rico, A., Illana, B., Salas, M. y Meijer, W. J. 2000. The putative coiled coil domain of the  $\Phi$ 29 terminal protein is a major determinant involved in recognition of the origin of replication. *J Biol Chem*, 275, 40529-40538.
- Serrano-Heras, G., Bravo, A. y Salas, M. 2008. Phage  $\Phi$ 29 protein p56 prevents viral DNA replication impairment caused by uracil excision activity of uracil-DNA glycosylase. *Proc Natl Acad Sci U S A*, 105, 19044-19049.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. y Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13, 2498-2504.

Sharma, S., Chatterjee, S., Datta, S., Prasad, R., Dubey, D., Prasad, R. K. y Vairale, M. G. 2016. Bacteriophages and its applications: an overview. *Folia Microbiol (Praha)*, DOI: 10.1007/s12223-016-0471-x.

Shcherbakova, P. V., Pavlov, Y. I., Chilkova, O., Rogozin, I. B., Johansson, E. y Kunkel, T. A. 2003. Unique error signature of the four-subunit yeast DNA polymerase  $\epsilon$ . *J Biol Chem*, 278, 43770-43780.

Sozhamannan, S., McKinstry, M., Lentz, S. M., Jalasvuori, M., McAfee, F., Smith, A., Dabbs, J., Ackermann, H. W., Bamford, J. K., Mateczun, A. y Read, T. D. 2008. Molecular characterization of a variant of *Bacillus anthracis*-specific phage AP50 with improved bacteriolytic activity. *Appl Environ Microbiol*, 74, 6792-6796.

St Charles, J. A., Liberti, S. E., Williams, J. S., Lujan, S. A. y Kunkel, T. A. 2015. Quantifying the contributions of base selectivity, proofreading and mismatch repair to nuclear DNA replication in *Saccharomyces cerevisiae*. *DNA Repair (Amst)*, 31, 41-51.

Steitz, T. A., Smerdon, S. J., Jager, J. y Joyce, C. M. 1994. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science*, 266, 2022-2025.

Stellberger, T., Hauser, R., Baiker, A., Pothineni, V. R., Haas, J. y Uetz, P. 2010. Improving the yeast two-hybrid system with permutated fusions proteins: the Varicella Zoster Virus interactome. *Proteome Sci*, 8, 8.

Strauss, B. S. 2002. The "A" rule revisited: polymerases as determinants of mutational specificity. *DNA Repair (Amst)*, 1, 125-135.

Stromsten, N. J., Benson, S. D., Burnett, R. M., Bamford, D. H. y Bamford, J. K. 2003a. The *Bacillus thuringiensis* linear double-stranded DNA phage Bam35, which is highly similar to the *Bacillus cereus* linear plasmid pBClin15, has a prophage state. *J Bacteriol*, 185, 6985-6989.

Stromsten, N. J., Bamford, D. H. y Bamford, J. K. 2003b. The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. *J Virol*, 77, 6314-6321.

Stromsten, N. J., Bamford, D. H. y Bamford, J. K. 2005. *In vitro* DNA packaging of PRD1: a common mechanism for internal-membrane viruses. *J Mol Biol*, 348, 617-629.

- Studier, F. W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, 41, 207-234.
- Suttle, C. A. 2007. Marine *viruses*--major players in the global ecosystem. *Nat Rev Microbiol*, 5, 801-812.
- Tabor, S. y Richardson, C. C. 1992. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. 1985. *Biotechnology*, 24, 280-284.
- Tanguy Le Gac, N., Delagoutte, E., Germain, M. y Villani, G. 2004. Inactivation of the 3'-5' exonuclease of the replicative T4 DNA polymerase allows translesion DNA synthesis at an abasic site. *J Mol Biol*, 336, 1023-1034.
- Uchiyama, Y., Takeuchi, R., Kodera, H. y Sakaguchi, K. 2009. Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie*, 91, 165-170.
- Uemori, T., Sato, Y., Kato, I., Doi, H. y Ishino, Y. 1997. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes Cells*, 2, 499-512.
- Verheust, C., Jensen, G. y Mahillon, J. 2003. pGIL01, a linear tectiviral plasmid prophage originating from *Bacillus thuringiensis* serovar israelensis. *Microbiology*, 149, 2083-2092.
- Verheust, C., Fornelos, N. y Mahillon, J. 2005. GIL16, a new gram-positive tectiviral phage related to the *Bacillus thuringiensis* GIL01 and the *Bacillus cereus* pBClin15 elements. *J Bacteriol*, 187, 1966-1973.
- Wang, F. y Yang, W. 2009. Structural insight into translesion synthesis by DNA Pol II. *Cell*, 139, 1279-1289.
- Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H. y Steitz, T. A. 1997. Crystal structure of a pol  $\alpha$  family replication DNA polymerase from bacteriophage RB69. *Cell*, 89, 1087-1099.
- Yamanaka, K., Minko, I. G., Takata, K., Kolbanovskiy, A., Kozekov, I. D., Wood, R. D., Rizzo, C. J. y Lloyd, R. S. 2010. Novel enzymatic function of DNA polymerase  $\nu$  in translesion DNA synthesis past major groove DNA-peptide and DNA-DNA cross-links. *Chem Res Toxicol*, 23, 689-695.

Yang, J. y Zhang, Y. 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*, 43, 174-181.

Yang, W. 2003. Damage repair DNA polymerases Y. *Curr Opin Struct Biol*, 13, 23-30.

Yang, W. 2005. Portraits of a Y-family DNA polymerase. *FEBS Lett*, 579, 868-872.

Yang, W. 2014. An overview of Y-Family DNA polymerases and a case study of human DNA polymerase  $\eta$ . *Biochemistry*, 53, 2793-2803.

Yoon, J. H., Roy Choudhury, J., Park, J., Prakash, S. y Prakash, L. 2014. A role for DNA polymerase  $\theta$  in promoting replication through oxidative DNA lesion, thymine glycol, in human cells. *J Biol Chem*, 289, 13177-13185.

Zaballos, A. y Salas, M. 1989. Functional domains in the bacteriophage  $\Phi$ 29 terminal protein for interaction with the  $\Phi$ 29 DNA polymerase and with DNA. *Nucleic Acids Res*, 17, 10353-10366.

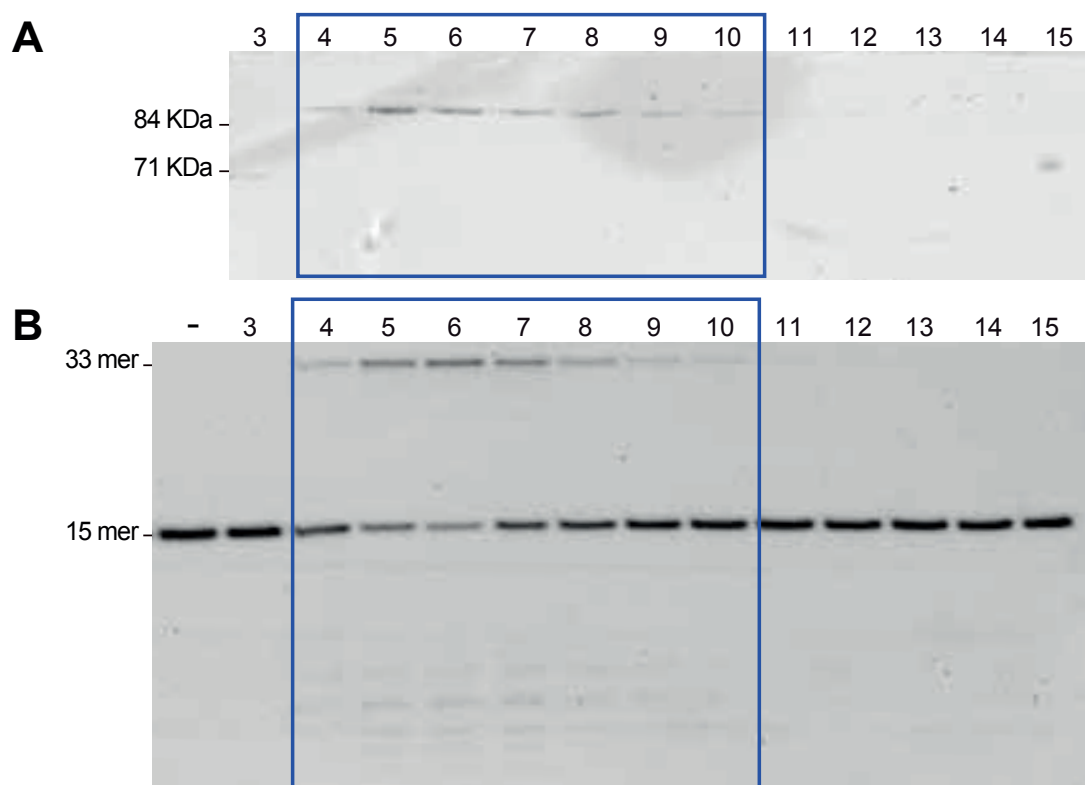
Zhou, B. L., Pata, J. D. y Steitz, T. A. 2001. Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol Cell*, 8, 427-437.

Zhu, Y., Song, L., Stroud, J. y Parris, D. S. 2008. Mechanisms by which herpes simplex virus DNA polymerase limits translesion synthesis through abasic sites. *DNA Repair (Amst)*, 7, 95-107.

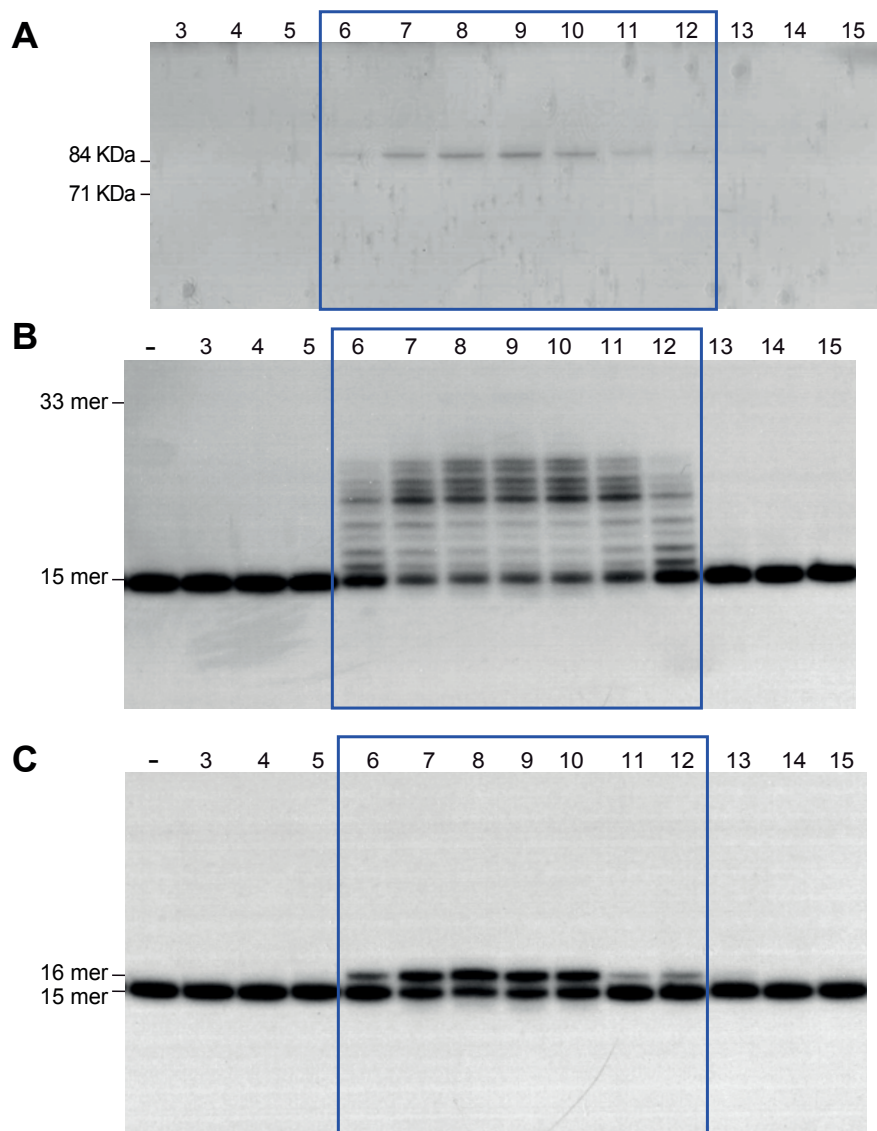








**Figura S1. Análisis de la replicación intrínseca de la B35DNAP mediante gradiente de glicerol.** A) Gel 12% SDS-PAGE teñido con azul de Coomasie de las fracciones del gradiente de glicerol de la B35DNAP. El gradiente se realizó con 8  $\mu$ g de la B35DNAP en un rango del 15 al 30% de glicerol con 20 mM de sulfato amónico y 80 mM de NaCl en tampón A sin glicerol (composición en página 42). Después de centrifugar a 58.000 rpm durante 24 horas a 4 °C se recogieron las diferentes fracciones para su posterior análisis. B) Ensayo de replicación del sustrato C15/M33GTT por la B35DNAP presente en las diferentes fracciones del gradiente. El ensayo se realizó como se describe en Materiales y Métodos 8.1 (página 49) pero con 2  $\mu$ L de cada fracción del gradiente y 1  $\mu$ M de dNTPs. La incubación se realizó a 30 °C durante 2 horas. El número de la fracción se indica en la parte superior de los geles, correspondiendo las fracciones más bajas al mayor porcentaje de glicerol. El cuadrado azul resalta las fracciones del gradiente con proteína y con actividad.



**Figura S2. Análisis de la replicación intrínseca de la variante  $\Delta$ TPR2 de la B35DNAP mediante gradiente de glicerol.** A) Gel 12% SDS-PAGE teñido con azul de Coomassie de las fracciones del gradiente de glicerol de la variante  $\Delta$ TPR2. El gradiente se realizó con 30  $\mu$ g de la  $\Delta$ TPR2 en un rango del 15 al 30% de glicerol con 20 mM de sulfato amónico y 80 mM de NaCl en tampón A sin glicerol (composición en página 42). Después de centrifugar a 58.000 rpm durante 24 horas a 4 °C se recogieron las diferentes fracciones para su posterior análisis. B) Ensayo de replicación del sustrato C15/M33GTT por la  $\Delta$ TPR2 presente en las diferentes fracciones del gradiente. El ensayo se realizó como se describe en Materiales y Métodos 8.1 (página 49) pero con 1  $\mu$ L de cada fracción del gradiente y 100  $\mu$ M de dNTPs. La incubación se realizó a 37 °C durante 30 min. C) Ensayo de replicación del sustrato C15/M33GFT por la  $\Delta$ TPR2 presente en las diferentes fracciones del gradiente. El ensayo se realizó en las mismas condiciones que en el panel B. El número de la fracción se indica en la parte superior de los geles, correspondiendo las fracciones más bajas al mayor porcentaje de glicerol. El cuadrado azul resalta las fracciones del gradiente con proteína y con actividad.